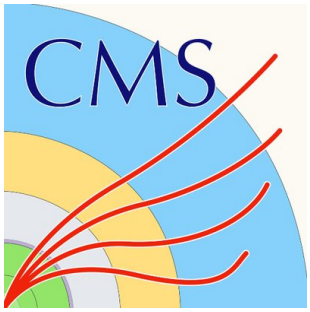


SEARCH FOR SUPERSYMMETRY AND THE MEASUREMENT OF THE TOP PAIR PRODUCTION IN ASSOCIATION WITH A W OR A Z BOSON WITH THE CMS DETECTOR AT THE LHC

KHVASTUNOV ILLIA

Supervisors: Prof. Dr. Didar Dobur and Dr. Maxim Titov
Proefschrift tot het bekomen van de graad van
Doctor in de Wetenschappen: Fysica
Academiejaar 2018 – 2019
Universiteit Gent

Faculteit Wetenschappen
Vakgroep Fysica en Sterrenkunde
May 2019

ABSTRACT

The Standard Model (SM) of particle physics is one of the most successful theories of the past century. It summarizes our understanding of all the known elementary particles and many predictions of this theory got confirmed over the past decades. The discovery of the BEH boson in 2012 by the CMS and ATLAS Collaborations was the culminating achievement of the SM. While the SM describes current experimental data very well, the indication of the dark matter content in the universe, the matter-antimatter asymmetry and neutrino oscillations are among the phenomena the SM cannot predict. Additionally, it should be mentioned that the SM covers only three of the four fundamental forces of nature, namely the electromagnetic, weak and strong interactions. Gravity, the dominant force at large scales, is simply absent in the SM. Even though the SM has succeeded in predicting and explaining quite a wide range of physics processes, it can not be the ultimate theory describing the universe.

The large data volume and an increase in the energy of the proton beams at the Large Hadron Collider in its second run had to answer if the supersymmetry (SUSY) exists in real life. The first part of this thesis is dedicated to the searches of strong SUSY production in the events with at least 3 electrons or muons, jets, and missing transverse momentum. No significant excess above the expected SM background is observed in the data set collected in 2016. Exclusion limits at 95% confidence level are computed for four different SUSY simplified models with pair production of gluinos or third-generation squarks. In the model with gluino pair production, with subsequent decays into a top quark-antiquark pair and a neutralino, gluinos with masses smaller than 1610 GeV are excluded for a massless lightest SUSY particle. In the case of bottom squark pair production, the bottom squark masses are excluded up to 840 GeV for charginos lighter than 200 GeV. For a simplified model of heavy top squark pair production, the $\widetilde{t}_2$ mass is excluded up to 720, 780, or 710 GeV for models with an exclusive $\widetilde{t}_2 \to \widetilde{t}_1 H$ decay, an exclusive $\widetilde{t}_2 \to \widetilde{t}_1 Z$ decay, or an equally probable mix of those two decays.

In the near future, no significant increase in energy of the proton beams and in the amount of collected data are anticipated, therefore to find physics beyond the SM we might need to reevaluate our strategy. One possibility is to consider a model-independent approach and to search for new physics in abundantly occurring electroweak-scale processes, which involve recently discovered heavy particles such as BEH, W and Z bosons and top quarks. The relatively high production rate of top pair in association with massive vector bosons ($t\bar{t}V$) allows us to study them in the first place. The second part of the thesis is dedicated to the inclusive cross section measurement for $t\bar{t}W$ and $t\bar{t}Z$ processes. The measurement of the $t\bar{t}W$ production is performed in the same-charge dilepton final state and reached a precision level of 20% with the data collected by the CMS experiment in 2016. The cornerstone of this measurement is the usage of a multivariate technique that helped to achieve the discovery level for the $t\bar{t}W$ process in $\sqrt{s} = 13$ TeV data set. The data sample used to measure precisely the $t\bar{t}Z$ production was collected by the CMS experiment during 2016 and 2017. The inclusive cross section is measured in the three- and four-lepton final states reaching the precision level of 8%. The reported result is the first experimental measurement that is more precise than the state-of-the-art theoretical calculations estimated at NLO QCD and electroweak accuracy for the $t\bar{t}Z$ production.

Dr. Hoenikker used to say that
any scientist who couldn't explain to
an eight-year-old what he was doing was a charlatan.

— Kurt Vonnegut, Cat's Cradle

ACKNOWLEDGMENTS

This thesis brings an end to four incredible years that gave me an opportunity to collaborate with many unforgettable people. Hereby I would like to express my gratitude to everybody who facilitated my work during the PhD research.

First of all, I would like to thank my supervisors Didar Dobur and Maxim Titov. Didar helped me in all aspects of my research and was always supportive both in constructive and criticizing ways. Maxim's doors were always open to answer my questions, or to give me an advice.

Doing a research in such a huge collaboration as CMS means to work in an international environment and share the experience with many brilliant people. I cannot thank Lesya Shchutska enough for giving me a helping hand in the very beginning of my research. I would like express my appreciation to all members of RA7 group for their guidance and advice during the preparation of both papers related to the SUSY search. I am grateful to the ttV team for helping me realize how important and valuable our work is. Presenting it at numerous conferences made me feel very proud for our research with Deniz Poyraz and Mirena Paneva.

Many thanks to Willem Verbeke and Tom Cornelis, which were not only the amazing office mates, but also helped in development of the ttZ/tZq Lepton MVA discrimination tool and in improving the systematic uncertainty of the $t\bar{t}Z$ cross section measurement. This was only possible thanks to the collaborative effort.

I would like to thank all members of the CMS group at the University of Ghent for providing a very friendly environment and also for tireless proofreading and helpful suggestions that shaped this thesis: my office mate Martina Vit, Marek Niedziela, Daniele Trocino, Basile Vermassen, Gianny Mestdach, Liam Wezenbeek, Tu Thong Tran.

Finally, thanks to my parents and friends for the support they gave me all these years. A special mention goes to my beloved wife Lili, for her unconditional support, love and encouragement.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

SM Standard Model

QED Quantum Electrodynamics

QCD Quantum Chromodynamics

QFT Quantum Field Theory

BEH Brout-Englert-Higgs

CERN Conseil européen pour la recherche nucléaire (European Organization for Nuclear Research)

LEP Large Electron–Positron Collider

SSB Spontaneous Symmetry Breaking

NP New Physics

EFT Effective Field Theory

GUT Grand Unified Theory

SUSY Supersymmetry

MSSM Minimal Supersymmetric Standard Model

LSP Lightest SUSY particle

SMS Simplified Model Spectra

LHC Large Hadron Collider

ATLAS A Toroidal LHC ApparatuS

CMS Compact Muon Solenoid

SLC Stanford Linear Collider

ALICE A Large Ion Collider Experiment

LHCb	Large Hadron Collider beauty
PU	Pile-up
ECAL	Electromagnetic calorimeter
HCAL	Hadron calorimeter
HLT	High Level Trigger
DT	Drift Tube
CSC	Cathode Strip Chamber
RPC	Resistive Plate Chamber
GEM	Gas Electron Multiplier
ME	Matrix Element
SM	Standard Model
BSM	Beyond the Standard Model
LO	Leading Order
NLO	Next-to-leading Order
ISR	Initial State Radiation
FSR	Final State Radiation
CKM	Cabibbo-Kobayashi- Maskawa
VEV	Vacuum Expectation Value
PDF	Parton Distribution Function
POG	Physics Object Group
MC	Monte Carlo
PF	Particle Flow
EA	Effective Area
ROC	Receiver Operating Characteristic
JEC	Jet Energy Correction
JER	Jet Energy Resolution
OCSF	Opposite Charge Same Flavor
SR	Signal Region
CL	Confidence Level
MVA	Multivariate Analysis
BDT	Boosted Decision Tree

1

INTRODUCTION

The Standard Model (SM) of particle physics is one of the most successful theories of the last century. It summarizes our understanding of all known elementary particles and forces; many predictions of this theory have been confirmed in precision experiments over the past decades. The discovery of the Brout-Englert-Higgs (BEH) boson in 2012 by the CMS and ATLAS Collaborations was the culminating achievement of the SM [1, 2]. While the SM is renormalizable and could be consistent theory valid up to possibly Planck scale which describes current experimental data quite well, there are several phenomena it cannot satisfactorily accommodate such as the non-baryonic dark matter, the origin of matter-antimatter asymmetry, and the non-zero neutrino masses.

Supersymmetry (SUSY) is often considered as an attractive solution to several outstanding problems of the SM. The large data volumes of proton-proton collisions at the unprecedented center-of-mass energy ($\sqrt{s}$) reached by the Large Hadron Collider (LHC) provide a unique opportunity to search for new physics in general. If SUSY is realized in nature near the weak scale, the LHC is bound to discover it. Nonetheless, so far analyzed LHC data shows no sign of new physics. Even though a large volume of data sets from LHC still needs to be analyzed, these searches are not expected to extend the discovery reach of the LHC significantly. Alternatively, other strategies might have to be exploited to find new physics. One possibility is to consider a model-independent approach and to search for new physics through precision measurements of processes at the electroweak scale, which involve heavy particles such as BEH, W and Z bosons and top quarks. **In this thesis we present searches for new physics exploring both methodologies; direct search for SUSY and precision measurements in the top quark sector.**

The analysis concerning search for SUSY used the data set from LHC run in 2015 and 2016 and led to stringent constraints on the masses of the SUSY particles such as gluinos, 3rd generation squarks, and lightest SUSY particles - neutralinos.

Using the data collected in 2016 and 2017 we performed cross section measurements for top pair production in association with massive vector bosons ($t\bar{t}W$, $t\bar{t}Z$). Several dedicated techniques were developed to increase the precision in the measurements. Most important ones are the improved electron and muon identification using machine learning techniques, which led to significantly increased signal purity, and precise determination of backgrounds due to misidentified leptons using data control samples. The results of this study present the most precise measurements of these processes available to date. In particular, the precision reached in $t\bar{t}Z$ is for the first time better than the state-of-the-art theoretical calculations estimated at next-to-leading order quantum chromodynamics and electroweak accuracy. The methodologies developed within the context of this analysis were also applied in other measurements performed by the UGent CMS group, such as the observation of single top production in association with a Z boson [3]. These measurements provide a solid ground for other precision studies at the electroweak scale, where the differential distributions can be studied in detail and can be

used in the context of Effective Field Theory (EFT) to constrain new physics at scales higher than the range probed with direct searches at the LHC.

In summary, this thesis gives a comprehensive overview of the studies performed between 2015 and 2019 using data collected by the CMS experiment in the proton-proton collision at $\sqrt{s} = 13$ TeV. The thesis starts with a theoretical overview of the SM in Chapter 2. Chapter 3 introduces the LHC and the CMS detector components including trigger and online processing. Chapter 4 describes Monte Carlo (MC) simulation models and programs used in this thesis for collision event simulation. Object identification and event reconstruction are presented in Chapter 5. Chapter 6 summarizes physics analysis techniques for the SUSY searches in events with multiple charged leptons, jets, and missing transverse momentum. These measurements have been published in the peer-reviewed CMS collaboration papers:

- CMS Collaboration, "Search for supersymmetry with multiple charged leptons in proton-proton collisions at $\sqrt{s} = 13$ TeV", **Eur. Phys. J. C** 77 (2017) 635
- CMS Collaboration, "Search for supersymmetry in events with at least three electrons or muons, jets, and missing transverse momentum in proton-proton collisions at $\sqrt{s} = 13$ TeV", **JHEP** 02 (2018) 067

Preliminary results of these studies are also published as a CMS public note:

- CMS Collaboration, "Search for SUSY with multileptons in 13 TeV data", Physics Analysis Summary, **CMS-SUS-**16-022.

Chapter 7 summarizes the measurement of top pair production in association with a W boson, using the 2016 dataset collected by the CMS experiment; the results are published:

- CMS Collaboration, "Measurement of the cross section for top quark pair production in association with a W or Z boson in proton-proton collisions at $\sqrt{s} = 13$ TeV", **JHEP** 08 (2018) 011

Preliminary results of this measurement are also available as a CMS public note:

- CMS Collaboration, "Measurement of the top pair-production in association with a W or Z boson in proton-proton collisions at 13 TeV", Physics Analysis Summary, **CMS-TOP-**16-017.

Chapter 8 presents the results of the cross section measurement of top pair production in association with a Z boson. The results of these measurement are in preparation to be submitted for publication. The preliminary results are published in a physics analysis summary by the CMS collaboration:

- CMS Collaboration, "Measurement of top quark pair production in association with a Z boson in proton-proton collisions at $\sqrt{s} = 13$ TeV", Physics Analysis Summary, **CMS-TOP-**18-009.

Finally, a summary and prospects for future analyses are given in Chapter 10.

The results obtained in this thesis were presented by the PhD candidate in various international conferences:

- "Search for SUSY with multileptons in proton-proton collisions in 13 TeV center-of-mass energy data using CMS detector", poster presented at the **General Scientific Meeting of the Belgian Physical Society**, May 2016, Ghent.

- "Studies of tt+V processes in the CMS Experiment", talk at the **Top Quark Physics workshop**, September 2016, Olomouc, Czech Republic. This preliminary result was presented the co-called ICHEP dataset, representing about 30% of the total statistics available by the end of the 2016. In addition, new measurement on the top pair production with a photon at 8 TeV center-of-mass energy was also reported.

- "Studies of the tt+X in CMS", talk at the **Large Hadron Collider Physics Conference**, May 2017, Shanghai, China. This talk outlined first CMS preliminary result on the measurement of top pair production with a vector boson, using the whole 2016 dataset. Additionally, the most recent results were presented for the measurement of various processes in top quark sector, among which are the top pair production with a photon, top pair production with a pair of b quarks and four top pair production.

- "Measurement of the top quark pair-production in association with a W or Z boson in pp collisions at 13 TeV with full 2016 dataset at CMS experiment", talk at the **La Thuile Conference**, March 2018, La Thuile, Italy. This talk was given during the Young Scientific Forum, where young researchers can deliver their results to the broad particle physics community.

2

THEORETICAL OVERVIEW

In the quantum mechanical description, subatomic particles are described by wave functions that evolve according to the Schrödinger equation. In a quantum field theory (QFT) the laws of quantum mechanics are applied to the fields [4], i.e. continuous substances, spread throughout space and time. One QFT paradigm, which exhibits a wide range of physics, is the Standard Model (SM). This model is the most successful theory describing the elementary particles, as well as three of the four fundamental forces in nature, namely the electromagnetic, weak and strong. Unfortunately, the fourth force - gravity, the dominant force at large scales,- is absent in this model. The ripples of a field are quantized into a bundle of energy, which can be thought of as particles within the QFT. For example, a cluster of the electromagnetic field can be considered as an electron.

An overview of the theoretical background for the work done in this thesis is given in this Chapter[1]. The main aspects and problems of the SM are discussed as well as a potential solution, such as Supersymmetry (SUSY), and an alternative approach to search for new physics phenomena.

2.1 THE STANDARD MODEL

Particles in the SM can be classified according to their intrinsic angular momentum (i.e. spin), which is related to how the wave function describing an ensemble of identical particles behaves under interchange of any two particles. Bosons, whose wave functions do not transform under such exchange, have integer values of spin, while fermions, whose wave functions change a sign under exchange of two particles with the same quantum numbers. The bosons obey Bose-Einstein statistics [6], while fermions obey Fermi-Dirac statistics [7]. There is a sacrosanct principle of QFT, called the spin-statistics theorem [8], which gives rise to the famous Pauli principle: two fermions can not occupy the same quantum state. No such restriction exists for bosons [9]. Fundamental particles can also be characterized by their chirality, the intrinsic property which specifies the direction in which the phase of the wave function rotates in the complex plane, i.e. clockwise is referred to as left-handed, and counterclockwise as right-handed. Additionally, fundamental particles are further characterized by their mass and by quantum numbers or charges that determine which of the fundamental forces they participate in.

Fermions are split into two groups: quarks and leptons. Quarks carry color charge, indicating that they participate in the strong interaction, while leptons do not. Quarks also carry an electric charge and, therefore, engage in the electromagnetic interaction. They can be either of an up-type, with an electric charge of $+\frac{2}{3}|e|$, or a down-type, with an electric charge of $-\frac{1}{3}|e|$. Leptons likewise can be electron-type or neutrino-type. Electrons carry an electric charge of $-|e|$ while neutrinos have no electric charge.

1 The summary of the SM given in Section 2.1 largely follows Ref. [5].

All fermions are grouped into three generations. The particles in second and third generations can be considered as massive copies of the first generation particles, though an exception can be made for the neutrinos, where the mass hierarchy is currently unknown. This hierarchy causes particles of higher generations to decay to the lightest one, which explains why matter around us is made of particles from the first generation. There is currently no evidence for the existence of a fourth generation [10]. One fundamental measurement was done by the ALEPH experiment, which ruled out additional neutrinos with a mass less than one half of the Z boson mass (around 45 GeV) by measurements of the Z boson branching ratio to hadrons [11]. Each generation is divided into two types of leptons (electron-like and neutrino-like) and two types of quarks (up-type and down-type), so in total there exist six flavors of quarks and six flavors of leptons. For each fermion, a corresponding antifermion with the same mass but with opposite quantum numbers exists. Charged massive fermions are described by Dirac spinor fields (spin equals $\frac{1}{2}$), which have a component for each of the two chirality states. The uncharged fermions, the neutrinos, carry only left-handed chirality, while anti-neutrinos carry only right-handed chirality.

The fundamental forces result from couplings between fundamental fields. For example, the force between two excitations of a fermionic field is mediated by an excitation of a fundamental bosonic field to which it is coupled, which can be considered as an interaction in which a boson is exchanged between two fermions. The electromagnetic interaction, responsible for binding electrons in atoms, is mediated through the exchange of an excitation of the electromagnetic field, a photon (γ), which is electrically neutral and massless. The strong force is responsible for holding quarks together to form hadrons and for binding protons and neutrons together to form atomic nuclei. It is mediated via the exchange of massless bosons called gluons. The weak interaction is the mechanism responsible for changing the particle flavor and the process known as a β decay. The mediators of weak interaction are the W and Z bosons, discovered at CERN in 1983 [12, 13]. The SM also includes a scalar boson (spin equals 0), the Higgs boson, which is the particle associated with the Brout-Englert-Higgs (BEH) field. This field is responsible for the mass generation for the W and Z bosons and fermions.

The summary of the particles, their properties and the fundamental forces which they participate in, is given in Fig. 2.1. In the next few subsections we discuss some relevant and basic aspects of the SM.

2.1.1 *The Lagrangian formulation and principle of stationary action*

The SM is relativistic and its equations of motion are Lorentz invariant, or in other words, they do not depend on the choice of inertial frame. The equations in QFT can be obtained from the action by means of Hamilton's principle. Mathematically, we can introduce the Lagrangian density $\mathcal{L}$ [15] (hereafter, simply the Lagrangian) as a functional that takes the configuration of one or few fields, $\phi(x)$ and their derivatives, $\partial_\mu\phi$ and gives a number. The field values always depend on the point in space-time, denoted as x[2]. The action S can then be constructed by integrating the Lagrangian over space-time:

$$S = \int \mathcal{L}(\phi, \partial_\mu\phi) d^4x \tag{2.1}$$

2 this will be omitted in equations to simplify the notation

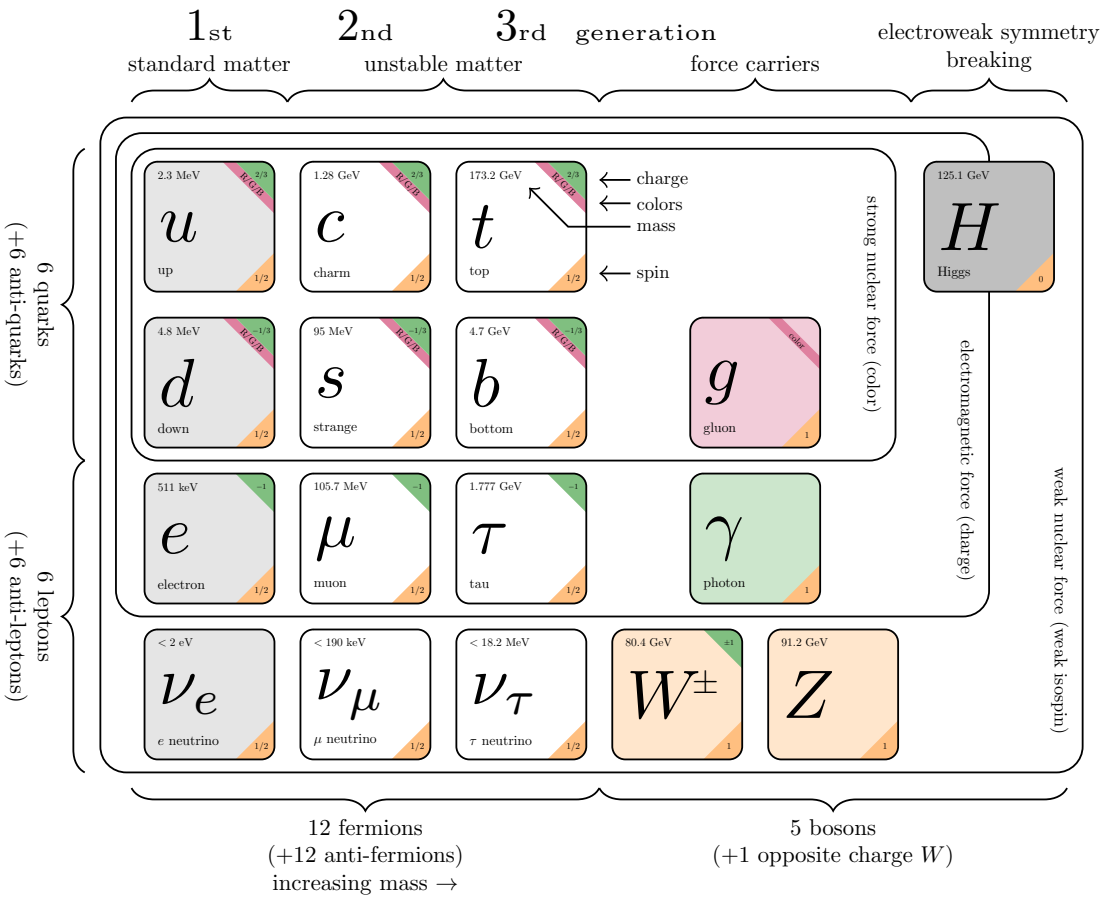


Figure 2.1: The summary of the SM of particle physics. The numbers are taken from [10]. The figure is made using the code taken from [14].

The action describes all possible trajectories for a system evolving from one given configuration to another. The principle of stationary action postulates that the true evolution of a physical system is a solution of the functional equation $\delta S = 0$.

2.1.2 *Symmetry and gauge invariance*

Symmetry, i.e. invariance under various transformations, plays an important role in the laws of nature. Symmetries can be either discrete or continuous. Examples of these two symmetries can be easily presented with geometrical figures. Rotation of a square is an example of a discrete transformation: only turns in 90, 180, 270 and obviously 360 degrees are symmetric. Rotation of a circle is an example of a continuous symmetry: after a rotation of any arbitrary angle, the circle looks the same as before.

A Lagrangian exhibits a continuous symmetry if it is invariant under a transformation of fields $\phi \to \phi\prime = U\phi$. If the continuous transformation U is constant at every point x in space-time it is said to be global. Noether's first theorem [16] states that a global continuous symmetry in the Lagrangian implies that an associated quantity is conserved. Three fundamental conservation laws in classic mechanics can be deduced from this theorem: invariance with respect to translations in time leads to the conservation of energy, invariance with respect to translations in space gives rise to the conservation of momentum, and invariance with respect to rotations in space gives rise to conservation

of angular momentum. In electrodynamics, the Lagrangian is invariant under global phase transformations, which gives rise to the conserved electric charge.

Field theories describe fields that can not be directly measured; instead, measurements are made of various observables such as energies, velocities, and charges. Different configurations, i.e. gauges, of the underlying fields may lead to the same set of observations. The components of the SM, i.e. quantum electromagnetic, weak and strong force theories, are examples of gauge theories. If the observable quantities of a physical theory do not change when a gauge transformation is performed (i.e., one possible gauge configuration is transformed to another), then the physical theory is said to be gauge invariant.

The SM is based on the requirement that the theory is invariant under certain types of local continuous gauge transformations $U(x)$, which unlike the global transformation U, can vary from point to point. The procedure used to enforce the symmetry on the fermionic fields necessitates the existence of additional gauge fields, i.e. bosonic fields, that are associated with the SM forces described earlier. To see how it works we first consider the U(1) group gauge symmetry and then continue with SU(2) and SU(3) groups. The unitary group U(n) is the group of unitary complex matrices with the size n × n; additional S in the name states for special, implying the determinant of matrices is equal to 1.

2.1.3 *Quantum Electrodynamics - abelian U(1) gauge theory*

Let's consider the Dirac Lagrangian for a spin-$\frac{1}{2}$ massive particle[3]:

$$\mathcal{L} = \bar{\psi}(i\gamma^{\mu}\partial_{\mu} - m)\psi, \tag{2.2}$$

where m and ψ are the mass and wave function associated with the particle, γ^{μ} are the Dirac matrices [4], $\bar{\psi}$ is the Hermitian conjugation of ψ and μ numerates the space-time coordinates. This Lagrangian is invariant under the global transformation of the U(1) group, which can represented as a turn in a complex plane [17]: $\psi \rightarrow \psi' = e^{iq\theta}\psi$. Here the charge, q, of the involved particle is written separately from the phase factor. This charge can be interpreted as a generator of the U(1) group. If we consider a local transformation $\psi \rightarrow \psi' = e^{iq\theta(x)}\psi$ which varies from point to point, the Lagrangian is then no longer invariant due to the derivative in the first term:

$$\partial_{\mu}(e^{iq\theta(x)}\psi) = iq\partial_{\mu}\theta(x)e^{iq\theta(x)}\psi + e^{iq\theta(x)}\partial_{\mu}\psi. \tag{2.3}$$

For the Lagrangian an additional term appears after the transformation:

$$\mathcal{L} \rightarrow \mathcal{L} - iq\bar{\psi}\gamma^{\mu}\partial_{\mu}\theta(x)\psi. \tag{2.4}$$

Clearly, the invariance has been spoiled, but a general procedure is well-known to recover the symmetry. The derivative in Eq. 2.2 is replaced with the covariant derivative

$$D_{\mu} = \partial_{\mu} - iqA_{\mu} \tag{2.5}$$

which includes an additional term whose role is to produce a cancellation of the term in Eq. 2.4 that spoiled the invariance. This procedure has resulted in the addition of a

3 For simplicity throughout this thesis natural units will be used in which $\hbar$=c=1

new vector field A_μ, transforming as $A_\mu \to A_\mu - \partial_\mu\theta$. As a new field has been added to the theory, a corresponding kinetic term must be added to the Lagrangian in order to describe the propagation of a spin-1 particle with a mass m [18]:

$$\mathcal{L}' = -\frac{1}{4}F^{\mu\nu}F_{\mu\nu} + m^2A^\mu A_\mu, \tag{2.6}$$

where

$$F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu. \tag{2.7}$$

Note that Lagrangian in 2.6 is not gauge invariant. The $F^{\mu\nu}$ is invariant under the gauge transformation:

$$\partial^\mu A^\nu - \partial^\nu A^\mu \to \partial^\mu(A^\nu - \partial^\nu\theta) - \partial^\nu(A^\mu - \partial^\mu\theta) \to \partial^\mu A^\nu - \partial^\nu A^\mu,$$

while the second term in Eq. 2.6 attains an additional term:

$$m^2A^\mu A_\mu \to m^2A^\mu A_\mu - 2m^2A^\mu\partial_\mu\theta.$$

Therefore, to leave the Lagrangian invariant for arbitrary angle $\theta(x)$, the mass should be equal to 0. Thus, starting from the globally invariant Dirac Lagrangian for fermions and requiring it to be locally invariant, a new massless vector field appeared as a matter of necessity. The A_μ can be interpreted as the photon field. Putting everything together we recognize this result as the quantum electrodynamics (QED) Lagrangian:

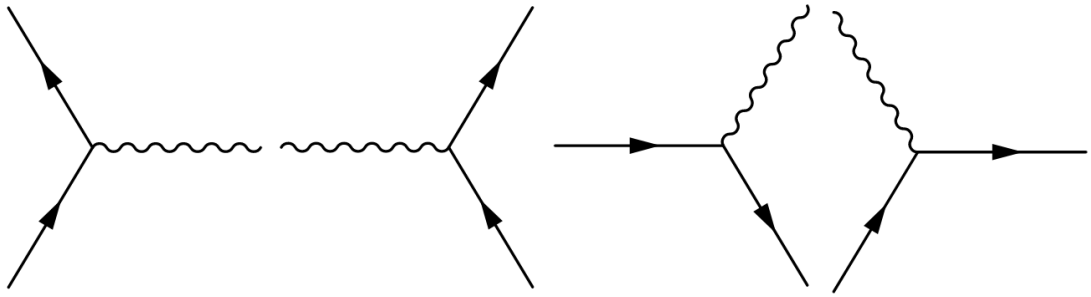
$$\mathcal{L}_{QED} = \bar{\psi}(i\gamma^\mu\partial_\mu - m)\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} + q\bar{\psi}\gamma^\mu\psi A_\mu. \tag{2.8}$$

The term in the new gauge invariant Lagrangian represents electromagnetic interactions of the fermions through their coupling with the photon field and the photons are the force carriers of these electromagnetic interactions. The strengths of these interactions are proportional to the charge q of the particles involved. QED is an example of an abelian gauge theory as successive applications of the phase commute with one another: the order in which they are applied can be reversed without changing the result.

2.1.4 *Renormalization*

The last term in the Eq. 2.8 can be rewritten as $j^\mu A_\mu$, where $j^\mu = q\bar{\psi}\gamma^\mu\psi$ is the conserved Noether current. This additional term, therefore, expresses the coupling with strength q between the conserved fermion current and the photon field. The resulting interaction vertexes, shown in Fig. 2.2, have therefore two fermion legs and one boson leg. These diagrams, showing lines for a certain class of particle paths through space-time, give a visible representation of quantum mechanical transition amplitudes from an initial to a final state and are called Feynman diagrams [19]. With each part of the diagram, Feynman rules can be associated, simplifying the derivation of a transition amplitude for the depicted process. The exact rules are shown in various books on the subject of particle physics [4, 9, 17] and will be skipped in this thesis. Besides being used to make

calculations more straightforward, Feynman diagrams can also simply be used to depict certain processes and interactions between particles. Matter particles are conventionally represented by straight lines with arrows pointing forward in time for fermions, and backward in time for antifermions. Electroweak gauge bosons tend to be represented by wavy lines, gluons by curly lines, and scalars by dashed lines.



(a) annihilation (b) pair creation (c) radiation (d) absorption

Figure 2.2: Feynman diagrams for interaction vertexes in a U(1) gauge theory, associated with the last term in the Eq. 2.8. In these diagrams, the arrow of time flows from left to right. Fermions are represented with solid lines with an arrow going forward in time, anti-fermions have their arrow of time going backwards in time. The gauge boson is represented with a wavy line. At each vertex, an incoming (ψ) and outgoing ($\bar{\psi}$) (anti-)fermion interact with a photon field leading to such physical processes as fermion and anti-fermion annihilation and pair creation, and photon radiation and absorption by the charged fermion [19].

The leading order for a fermion scattering process is shown in Fig. 2.3, additionally to the high-order diagrams, which can involve closed loops. For each additional vertex in the Feynman diagram, a factor $\alpha \sim e^2$ (e is the charge of an electron) is introduced in the scattering amplitude, while the virtual particles (i.e. the particles described by the internal lines in the Feynman diagram) contribute to the total scattering amplitude through their propagators which depend on the momentum of virtual particles. One needs to consider all possible momenta in the propagators, introducing integrals over momentum space. Since there is no limit on the range of the integration and on the number of diagrams, the probability may a priori diverge to infinity, which can be handled through renormalization. A theory with such property, i.e. a theory for which the series of the contributions from all diagrams converges is said to be renormalizable.

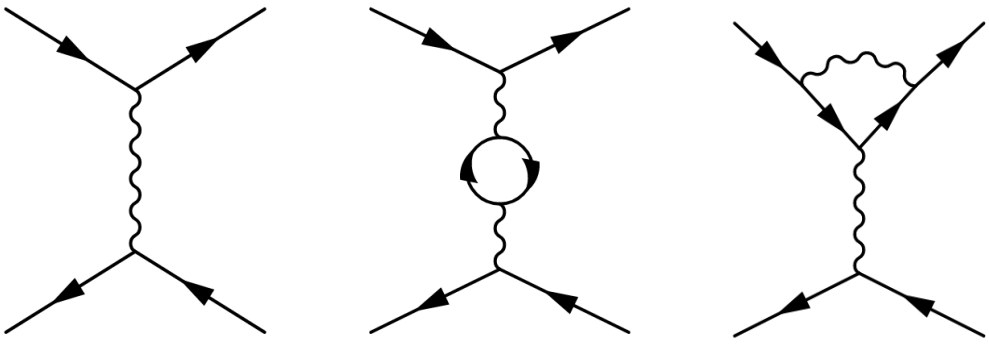


Figure 2.3: Feynman diagrams for a fermion scattering process. In addition to the first-order diagram (left) in which the incoming fermions are simply scattered by exchange of a gauge boson, two higher-order loop contributions (center and right) are shown [19].

The renormalization procedure hides the divergences by absorbing them in a redefinition of the coupling constant and the mass of the fermion. Because the higher-order contributions depend on the total momentum transfer, renormalization introduces a renormalization scale μ, the point at which the divergences canceling is performed. As a result, the renormalization coupling also depends on μ^2 according to:

$$\alpha(q^2) \approx \alpha(\mu^2) \frac{1}{1 - \frac{\alpha(\mu^2)}{3\pi} \log \frac{q^2}{\mu^2}}, \tag{2.9}$$

where q is the transfered momentum between the fermions[4] [18]. The electromagnetic coupling can be obtained by an appropriate renormalization of the electron charge defined at an arbitrary scale μ^2. From the Eq. 2.9 it could be seen that the the electromagnetic coupling parameter increases with an increase in transfered momentum. In the literature this effect is referred to as the "running of the coupling constant". At momentum transfers close to electron mass α approximately equals to $\frac{1}{137}$, while close to Z boson mass it is close to $\frac{1}{128}$. At the Large Electron–Positron Collider (LEP) it was experimentally confirmed that the strength of the electromagnetic interaction between two charged particles increases as the center-of-mass energy of the collision increases [20].

2.1.5 *Non-abelian gauge theories*

To make further progress and describe weak and strong interaction a Lagrangian invariance under local transformations of non-abelian groups, such as SU(2) and SU(3), must be considered. The non-abelian groups are the groups for which the generators of the group, T^a, do not commute:

$$[T^a, T^b] = i f^{abc} T^c,$$

where f^{abc} are called the structure constants [21]. A simple example of a non-abelian group is the rotation group SO(3) in three dimensions: the rotation in 90 degrees along one axis and then 90 degrees along a different axis doesn't give the same result as doing the rotation the other way round. In the following subsections, the electroweak and quantum chromodynamics (QCD) are discussed separately. For each of SU(2) and SU(3) group, a representation t^a is first chosen for the generators of the group similarly to q in U(1) group. Next step, we demand that the theory be invariant under a local SU(n) transformation, $\psi \to e^{ig\alpha^a(x)t^a}\psi$, where $\alpha^a(x)$ are arbitrary functions of x and g parameterizes the strength of the interaction. As was introduced previously, the derivative will be replaced by a covariant derivative:

$$D_\mu = \partial_\mu - igA^a_\mu t^a.$$

In this term a new vector field A^a_μ appears. The field tensor then should be extended with respect to Eq. 2.7:

$$F^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + g f^{abc} A^b_\mu A^c_\nu.$$

4 This is a usual definition, please do not mix with a charge defined earlier

This extension happens as a result of the non-commutative nature of the generators of SU(n) group [4]. The last term can be interpreted as the gauge boson self-interactions which do not appear in abelian theories such as QED. Note also that a gauge field has emerged corresponding to each generator of the group. The symmetry group used in the case of QED is U(1), which has a single generator and a single gauge field. In addition, Noether's theorem results in a single conserved charge. SU(n) has $n^2 - 1$ generators; therefore, an SU(n) gauge theory has $n^2 - 1$ gauge fields, and application of Noether's theorem yields n conserved charges.

The SM is based on the product of two special unitary groups and one unitary group: SU(3)×SU(2)×U(1). The SU(3) group describes the strong interactions, while the SU(2)×U(1) group describes both the weak and electromagnetic interactions.

2.1.6 *Quantum chromodynamics*

QCD, which provides a successful description of the strong interaction, emerges when the Lagrangian is required be invariant under local transformations of the non-abelian gauge group $\mathrm{SU(3)_c}$. Because it is SU(n) with $n = 3$, there are three conserved charges, referred to as color charges red, green, and blue. Similarly, there are $n^2 - 1 = 8$ generators of the group corresponding to the eight gauge fields, called gluons. Leptons, which carry no color charge, do not interact via the strong force. Each quark flavor can be expressed as a color triplet:

$$q = \begin{bmatrix} q_r \\ q_g \\ q_b \end{bmatrix}$$

where the q_x are Dirac spinors and the subscripts label the color states. The QCD Lagrangian is

$$\mathcal{L}^{QCD} = \mathcal{L}^{quarks} + \mathcal{L}^{gluons} = \sum_f \bar{q}_i^f (i\gamma_\mu D_{ij}^\mu - m_f \delta_{ij}) q_f^j - \frac{1}{4} \sum_{a=1}^{8} G_{\mu\nu}^a G_a^{\mu\nu},$$

where

$$D_{ij}^\mu = \partial_\mu \delta_{ij} - i g_s t_{ij}^a G_a^\mu,$$

t^a are the generators of SU(3) and can be represented as Gell-Mann matrices [18], g_s is the strong coupling constant, δ_{ij} - Kronecker delta (the i and j are color indexes running from 1 to 3). The gluon field strength tensor with color a is defined as

$$\mathbf{G}_{\mu\nu}^a = \partial_\mu G_\nu^a - \partial_\nu G_\mu^a + g_s f^{abc} G_\mu^b G_\nu^c. \tag{2.10}$$

Here and f^{abc} are defined by the commutation relation for the Gell-Mann matrices. As carriers of color charge, gluons can couple to one another; this "self-interaction" arises from the last term in Eq. 2.10.

2.1.7 *Electroweak unification*

The models that make an attempt to describe electroweak interaction (four fermion interaction theories like Fermi model) violate unitarity at high energy and are not renormalizable. In analogy with QED the solution of the problem emerges from construction of a gauge theory of weak interactions leading to the introduction of intermediate vector bosons with spin equal to 1. However, to model the short range of the weak interactions, such bosons could not have zero mass, and this term in the Lagrangian as we saw before would violate the gauge symmetry. The problem was solved (discussed in the next subsection) by the introduction of spontaneously broken symmetries (SSB), which then led to the prediction of the existence of the so-called BEH boson.

A unified description of the weak and electromagnetic forces was proposed by Glashow [22], Salam [23], and Weinberg [24], and it results from the assumption of invariance under an SU(2)×U(1) gauge symmetry. The SU(2) symmetry gives rise to the conserved weak isospin charge with three corresponding massless bosons (W^1, W^2, and W^3), and the U(1) symmetry gives rise to the conserved weak hypercharge and its associated massless B boson.

A major difference between the weak force and the other fundamental forces is that the weak force is a chiral theory, which does not have symmetry for different particle chiralities. To specify the field content it is necessary to split it into left- and right-handed components, which can be accomplished via the projection operators:

$$\psi_L = \frac{1}{2}(1-\gamma^5)\psi,$$

$$\psi_R = \frac{1}{2}(1+\gamma^5)\psi,$$

where γ^5 is the fifth γ matrix defined through other four γ matrices. The right-handed fermions do not participate in the weak interaction. For this reason, the weak isospin group is introduced, which is referred to as $SU(2)_L$. Mathematically, this means that right-handed fermions transform as isospin singlets and left-handed fermions transform as isospin doublets:

$$Q_L^i = \begin{bmatrix} u_L^i \\ d_L^i \end{bmatrix}, E_L^i = \begin{bmatrix} \nu_L^i \\ \ell_L^i \end{bmatrix}$$

$$f_R^i = e_R^i, u_R^i, d_R^i$$

where i runs over the three fermion generations. The mass term can be expressed in the following way:

$$-m_f\bar{\psi}\psi = -m_f(\bar{\psi}_R\psi_L + \bar{\psi}_L\psi_R)$$

Other two terms vanish because of the γ^5 matrix properties:

$$(1-\gamma^5)(1+\gamma^5) = 1-(\gamma^5)^2 = 0$$

The mass term as before is not gauge invariant because ψ_L changes under isospin rotations while ψ_R does not. The covariant derivative D_μ is

$$D_\mu = \partial_\mu - ig_1\frac{1}{2}W_\mu^a\sigma^a - ig_2\frac{Y}{2}B_\mu$$

where the a index runs from 1 to 3 over the generators of SU(2). Here σ^a are Pauli matrices, B_μ = U(1) gauge field, W_μ^a are 3 SU(2) gauge fields, g_1 and g_2 are SU(2) and U(1) coupling constants. Interactions between the two flavors in a doublet are mediated by the W^1 and W^2 bosons, which carry isospin $T_3 = \pm 1$, while the W^3 couples with two fermions of the same flavor has no weak isospin. The two flavors in the doublet have $T_3 = \frac{1}{2}$ and $T_3 = -\frac{1}{2}$ respectively. The right-handed particles carry no weak isospin charge, and do not interact with the $\mathrm{SU}(2)_L$ gauge bosons. The electromagnetic charge is retrieved as a simple sum out of the weak isospin and hypercharge using the Gell-Mann-Nishijima formula [25–27]:

$$Q = T_3 + \frac{Y}{2}$$

The term describing how fermions interact with the W_μ^a and B_μ fields is

$$\mathcal{L}_{int} = \sum \bar{f}^i(i\gamma^\mu D_\mu)f^i$$

where i = 1, 2, 3 refers to the three fermion generations, and f - all possible left and right fermions.

The unification of the forces leads to the mixing of the three $\mathrm{SU}(2)_L$ and the $\mathrm{U}(1)_Y$ gauge bosons, which form the observed W boson, Z boson and photon fields through the following linear combinations:

$$W_\mu^\pm = \frac{W_\mu^1 + iW_\mu^2}{\sqrt{2}}$$

$$Z_\mu = \cos\theta_W W_\mu^3 - \sin\theta_W B_\mu$$

$$A_\mu = \sin\theta_W W_\mu^3 + \cos\theta_W B_\mu$$

The Weinberg angle θ_W defines the mixing between the neutral $\mathrm{SU}(2)_L$ gauge field and the $\mathrm{U}(1)_Y$ gauge field. As a result, it links the coupling constants of the two symmetry groups:

$$\left.\begin{aligned}\sin\theta_W &= \frac{g_2}{\sqrt{g_1^2+g_2^2}}\\ \cos\theta_W &= \frac{g_1}{\sqrt{g_1^2+g_2^2}}\end{aligned}\right\} \rightarrow \tan\theta_W = \frac{g_2}{g_1} \tag{2.11}$$

This completes the description of an electroweak sector, which has two obvious drawbacks. As it was shown, there are four conserved charges, while the observed electroweak sector conserves only electric charge and the neutral component of weak isospin. Needless to say, that similarly to U(1) none of the SM fermions that participate in weak interaction can have mass.

2.1.8 *Electroweak symmetry breaking*

A model to dynamically generate mass for SM fermions and W and Z bosons was proposed independently by Brout and Englert [28], Higgs [29] and Guralnik, Hagen, and Kibble [30]. The gauge symmetry must somehow be broken, which occurs through a mechanism called SSB. SSB is "spontaneous" in the sense that "no external agency is responsible" for it [5], as occurs for example when gravity breaks the symmetry, making up and down different from left and right.

Let's consider a doublet of complex scalar fields for which the ground state does not respect the $SU(2)_L \times U(1)_Y$ symmetry:

$$\phi = \begin{bmatrix} \phi^+ \\ \phi^0 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{bmatrix}$$

The isospin doublet has weak hypercharge Y = 1, and thus the upper component with $T_3 = \frac{1}{2}$ has electric charge +1 while the $T_3 = -\frac{1}{2}$ component has no electric charge. The Lagrangian describing this weak isospin doublet contains a kinetic and potential term:

$$\mathcal{L} = \frac{1}{2}(\partial_\mu \phi)^\dagger (\partial^\mu \phi) - V(\phi), \tag{2.12}$$
$$V(\phi) = \mu^2 \phi^2 + \lambda \phi^4.$$

The Lagrangian contains a mass term μ^2 and a parameter λ describing the strength of the scalar field's self-interaction. To ensure a global minimum for the potential term, $V(\phi)$, λ is required to be positive. The mass term μ^2 can take negative or positive values. If μ^2 is positive, the global minimum is at $\phi = 0$, and the symmetry is respected. But in the case of $\mu^2 < 0$, the potential has no longer a minimum at $\phi = 0$, which is now a local maximum. Instead, we have a degenerate vacuum, described by

$$\phi^2_{min} = \frac{\mu^2}{2\lambda} = \frac{v^2}{2}$$

Here the variable v is introduced, which is used to denote so-called vacuum expectation value (VEV) for a Higgs field. Using the Goldstone theorem, we can choose a specific gauge for $SU(2)_L \times U(1)_Y$ where $\phi^+ = 0$ and fix neutral component of ϕ to neutral number. Specifically, the ground state, invariant under $U(1)_{EM}$ transformation looks like:

$$\phi_{ground} = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 \\ v \end{bmatrix}$$

with excited states defined with respect to this ground state in terms of a real field $h(x)$:

$$\phi = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 \\ v + h(x) \end{bmatrix} \tag{2.13}$$

After substituting Eq. 2.13 in the Lagrangian in Eq. 2.12 the terms which corresponds to massive $W^\pm$ and Z bosons will appear from $((\partial_\mu \phi)(\partial^\mu \phi))$ term. The $W^\pm$ and Z bosons attain the mass equal to $\frac{gv}{2}$ and $\frac{\sqrt{g'^2+g^2}v}{2}$, respectively. After the symmetry breaking, the electric charge is still conserved, and the photon is still massless. Note that the

U(1) symmetry that is respected is not the same U(1) symmetry that was present before SSB because the photon field is a mixture of fields from the spontaneously broken $SU(2)_L \times U(1)_Y$ symmetry.

The quarks and leptons are "born" massless, as the terms which contain $\bar{\psi}\psi$ are not invariant under the electroweak $U(2)_L \times U(1)_Y$ symmetry. However, it is still possible to write a Yukawa Lagrangian which describes the coupling between the fermions and the scalar Higgs field ϕ: $-\alpha_f \bar{\psi}\psi\phi$. When SSB happens this term splits into 2 pieces, one of which is a Yukawa coupling to the physical Higgs field ($h(x)$) and second is the pure fermion mass term, which contains VEV as a coefficient. Unfortunately, this splitting doesn't help us to calculate the particle masses - it simply trades one unknown parameter (mass of the fermion) for another (Yukawa coupling to the scalar field).

Searches for the BEH boson were performed by both ATLAS and CMS Collaborations with the data set collected in Run I. Five decay modes, namely $\gamma\gamma$, $ZZ \to 4\ell$, W^+W^-, $\tau^+\tau^-$ and $b\bar{b}$, were examined and the most significant excesses were observed in the decay modes $\gamma\gamma$ and $ZZ \to 4\ell$ at an invariant mass of the reconstructed particles around 125 GeV. In 2012 ATLAS and CMS Collaborations claimed the discovery of the BEH boson [1, 2]. The analyses of data set collected at Run II between 2015 and 2018 confirmed the observation and the properties of the BEH boson were studied, in particular in one of the most prominent final states with low expected background, namely the four charged lepton final state; the reconstructed invariant mass of four charged lepton with the data set collected by the CMS experiment in Run II (137.1 fb^{-1}) is shown in Fig. 2.4. The measured mass of the boson is equal to 125.26 $\pm$ 0.21 GeV [31], while the boson width is constrained to be $3.2^{+2.8}_{-2.2}$ MeV [32].

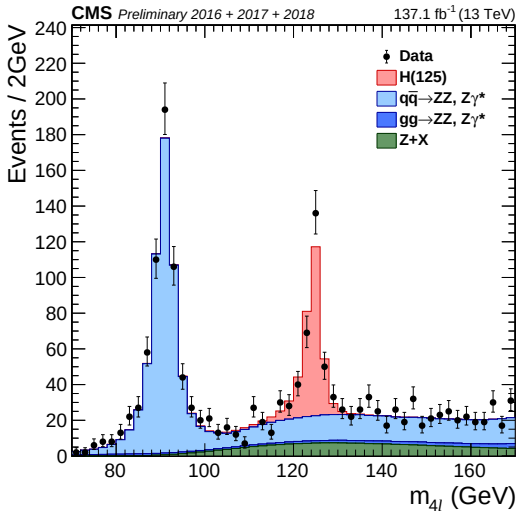


Figure 2.4: Distribution of the reconstructed four-lepton invariant mass. Points with error bars represent the data and stacked histograms represent the expected signal and background distributions. The BEH boson with a mass around 125 GeV is denoted as H(125). The figure is taken from [33].

Up to date the BEH boson has been examined in various decay modes: besides the discovery four-lepton and $\gamma\gamma$ channels, the observation of the BEH boson with five standard deviations was confirmed in the W^+W^-, $\tau^+\tau^-$ [34] and $b\bar{b}$ [35] final states separately. In addition, the BEH boson production in association with a top quark-antiquark pair was observed [36] and first searches for the BEH boson were performed in the $\mu\mu$ channel [37]. In all these channels, the Yukawa couplings to fermions and

coupling to vector bosons were measured by the CMS experiment with the 35.9 fb^{-1} of data collected in 2016 [38]. These measurements are illustrated as a function of the particle mass in Fig. 2.5. The fit to the measured coupling constants describes the SM prediction within one standard deviation, except for high masses close to the t quark mass. These precision measurements show that the observed boson is indeed the BEH scalar boson predicted in the SM.

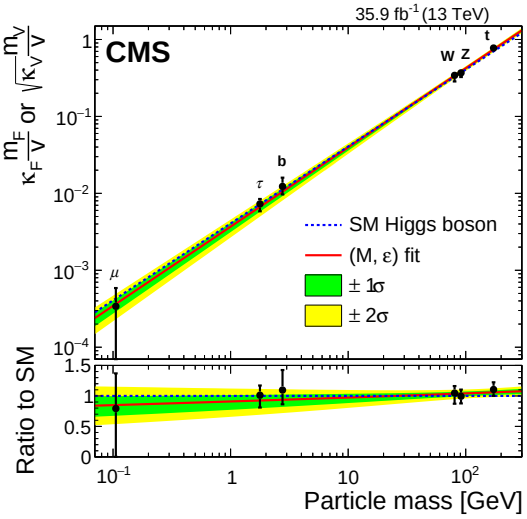


Figure 2.5: Best fit values of the Yukawa couplings to fermions and to vector bosons as a function of the particle mass obtained in the $\sqrt{s} = 13$ TeV data set collected using the CMS detector. The m_F and m_V correspond to fermion and vector boson masses, respectively, and v is the VEV. The dashed line indicates the predicted dependence on the particle mass in the case of the SM BEH boson. The solid red line indicates the best fit result with the corresponding 68% and 95% CL bands. The figure is taken from [38].

2.2 PHYSICS BEYOND THE STANDARD MODEL

The SM has been very successful. Plenty of particles and their properties were predicted before they were confirmed experimentally to exist, among them are the BEH, W and Z bosons. Though the theoretical work on the SM largely was completed in the 1970s, the most recent observation of the BEH boson in 2012 [1, 2] confirmed the mechanism of spontaneous electroweak symmetry breaking. Even though the SM has succeeded in predicting and explaining quite a wide range of physics processes, it can not be the ultimate theory describing the universe. The open questions to the SM roughly can be divided into two categories: first comprises experimental results not covered by the SM and second covers the theoretical problems of the SM which don't have fundamental explanation.

Cosmological observations have revealed that the SM particle spectrum only corresponds to 5% of the energy and matter content in the universe, while the remaining 27% and 68% are composed of so-called dark matter and dark energy [39]. Dark matter is a hypothetical type of matter that interacts with ordinary matter so weakly that it has not yet been observed directly, despite its gravitational influence has been indirectly observed in a wide range of astronomical and cosmological data [40]. The SM neutrinos are similar to dark matter, but they are not massive enough to constitute a significant

fraction of dark matter. Dark energy is an unknown type of energy associated with the vacuum of space, and it is responsible for the observed acceleration of the universe expansion. Furthermore, the SM cannot explain the imbalance between ordinary matter and antimatter known as the matter-antimatter asymmetry [41].

According to the SM, neutrinos are massless particles. However, neutrino oscillation experiments have shown that neutrinos are massive. Mass terms for the neutrinos can be added to the SM by hand, but this leads to a theoretical problem, i.e. the mass terms need to be extraordinarily small. The mass range in the SM is already 5 orders of magnitude wide, from electron mass of 0.511 MeV to top quark of 173 GeV. No explanation exists to tell us why the range, i.e. coupling to the BEH boson according to the spontaneous electroweak symmetry breaking mechanism, is so wide. Moreover, we know that at much larger scales, i.e. the Planck scale of 10^{19} GeV, the SM can no longer be a complete theory, as quantum gravity effects will then need to be included. This is usually referred to as the hierarchy problem.

Another theoretical problem is the number of free parameters in the SM, which are 19 numbers. The values of the parameters are known from experiments, but the origin of the values has no suitable explanation. Gravity, as already mentioned, is not included in the SM. Although gravity is sometimes considered too weak to be relevant in experimental physics, it governs the physics on a large scale. Gravity is well described under relativity; however, it is not predicted by the SM, and a unified theory that encompasses both quantum phenomena and gravity has not been formulated yet.

All mentioned problems point physicists to the fact that the theories beyond the standard model (BSM), or so-called new physics (NP), are necessary.

2.3 NEW PHYSICS IN THE TOP QUARK SECTOR

The top quark, the heaviest fundamental particle of the SM, was discovered quite recently in 1995 at Fermilab by the CDF and D0 collaborations [42, 43]. Its discovery was expected, because it was the only missing constituent in the third generation of fermions. At the LHC, top quarks are produced predominantly in pairs through gluon fusion and quark-antiquark annihilation.

Various theoretical motivations justify paying special attention to processes involving top quarks. Owing to its large mass, the top quark has a very short lifetime of around 5×10^{-25} s. Consequently, it decays before it has time to interact with any other particles, and it does not form bound states as other quarks do. This gives us a possibility to study a "bare" quark. Furthermore, because of its high mass, the top quarks contribution to the higher-loop diagrams plays a significant role, making the mass of top quark a crucial parameter in theory. The Yukawa coupling of the Higgs to the top quark is at the order of unity and is much larger than the next largest Yukawa coupling (see Fig. 2.5). This suggests that the top quark may have a special relationship to the Higgs field and can give us a hint to NP.

One of the most promising strategies to detect NP is by measuring the top quark coupling to gauge bosons and looking for deviations from the SM predictions. Most relevant for the thesis presented here are the processes of top quark pair production ($t\bar{t}$) in association with a W ($t\bar{t}W$) and a Z ($t\bar{t}Z$) bosons which became accessible with huge data sets collected so far. The leading order Feynman diagrams for $t\bar{t}W$ and $t\bar{t}Z$ processes are shown in Fig. 2.6. The $t\bar{t}Z$ process allows us to study the coupling between the top quark and the Z boson. This coupling is an important missing piece of information in the way of understanding the electroweak symmetry breaking mechanism. In principle,

indirect constraints on the top quark and a Z boson coupling can be obtained from the data collected at LEP. But the most precise constrain on the coupling can only be obtained from the direct measurement of the $t\bar{t}Z$ process [44]. The precision measurement of the $t\bar{t}Z$ inclusive cross section is presented in this thesis in Chapter 8.

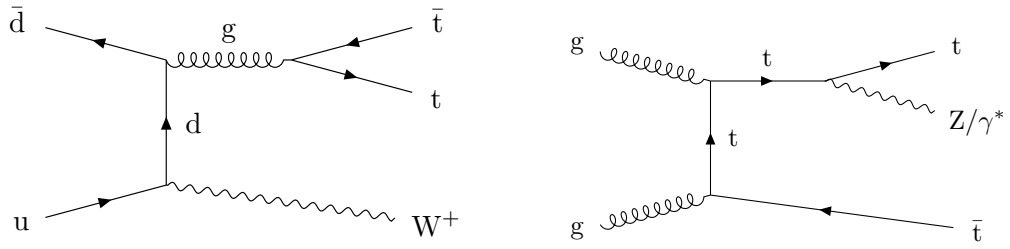


Figure 2.6: Dominant leading order Feynman diagrams for the $t\bar{t}W$ (left) and $t\bar{t}Z$ (right) processes at the LHC. The plots are taken from [45].

Besides the discovery of the top quark, Tevatron has facilitated precision measurements of the top quark properties. In most cases, the results were compatible with the predictions of the SM. However, a notable exception is represented by the forward-backward asymmetries, often also referred to as charge asymmetries, in top quark pair production. In short, this asymmetry can be defined as:

$$A_{FB} = \frac{N_F - N_B}{N_F + N_B}, \tag{2.14}$$

where N_F is the number of events in which some particular final state particle is moving "forward" with respect to some chosen direction, while N_B is the number of events with the final state particle moving "backwards".

Unfortunately, the asymmetry predicted by the SM at the LHC is much smaller than similar quantity at the Tevatron. The main reason is that the asymmetry is induced only by the fraction of top quark pair final states generated by quark-antiquark collisions, which at the LHC represent only around 15% of the total rate, compared to 85% at Tevatron. The smallness of the effect at the LHC makes it difficult to reach the sensitivity needed to measure the charge asymmetry in $t\bar{t}$ events, and probe possible deviations from the SM expectation. The latest measurements of top charge asymmetry performed so far using CMS and ATLAS data show a consistency with the SM predictions, though the best-measured value has a total uncertainty of 60% [46].

The $t\bar{t}W$ production is an alternative process for probing SM charge asymmetry in top quark pair production. At leading order (LO) production in QCD (see Fig. 2.6, left), it can only occur via a quark-antiquark annihilation, and no contribution from gluons in the initial state is possible. The small contribution from the symmetric gluon-gluon production channel, which only opens up at next-to-next-to-leading order, makes the resulting asymmetry significantly larger than in $t\bar{t}$ production [47]. The first step towards the charge asymmetry measurement in the $t\bar{t}W$ process is the $t\bar{t}W$ inclusive cross section measurement, which is presented in this thesis in the Chapter 7.

Moreover, both the $t\bar{t}W$ and $t\bar{t}Z$ cross section measurements can be interpreted in the context of an effective field theory (EFT) to constrain the Wilson coefficients [48]. In this way, cross section measurements can be used for NP searches in a model independent way.

In effective field theory approach, the SM Lagrangian is extended through higher order operators. To study the effects of NP, the expected cross sections are calculated as a function of the Wilson coefficients which parametrize the strength of the NP interactions. The details and most recent experimental constraints from $t\bar{t}W$ and $t\bar{t}Z$ measurements are reported in [45, 49].

2.4 SUPERSYMMETRY

As we already saw previously, symmetries are playing a crucial role in the SM. In 1974, Wess and Zumino introduced[5] a symmetry that mixes together fermions and bosons. Invariance under the space-time transformation of fermions into bosons and vice versa is called supersymmetry, in short SUSY. In principle, it is not so hard to construct the Lagrangian for such a symmetry, as long as the boson and fermion partners have the same mass. In reality, no such particles were observed experimentally. It means that this symmetry, if it exists of course, must be a broken symmetry. This way we can presume that the SUSY particles are much heavier than the accessible energy range at modern colliding machines.

There exist several theoretical justifications why this symmetry might exist. An introduction of a number of new particles modifies the energy dependence of the strong, weak and electromagnetic running coupling constants. Within a Grand Unified Theory (GUT) framework all three coupling constants appear to unify at around 10^{16} GeV if we include contribution from SUSY. Moreover, an important step can be made of including gravity in the overall unified framework.

SUSY also gives a solution to a hierarchy problem, which can be reformulated in the following way. We know that the BEH boson strongly couples to the most massive particles, in particular, coupling to the top quark is the largest. The SM is still required to be a renormalizable theory, which means that the BEH boson mass as an observed quantity should get a contribution from various loop diagrams involving in majority the contribution from the top quark. If a suggestion is made that the SM is valid not only in electroweak scale, but also can be extended far away to the Planck scale, then quantum corrections to the BEH boson mass will be proportional to the scale itself. It also means that SM Lagrangian will diverge quadratically at the Planck scale. Introducing two complex scalars will rule out this divergence because according to spin-statistics theorem [8] the contribution from these two scalars must be opposite to the contribution from the top quark.

Within minimal supersymmetry SM (MSSM) theory a discrete, multiplicative symmetry, called R-parity, could be introduced. Its quantum number is given by the formula:

$$R = (-1)^{3B+L+2s},$$

where B, L and s are the baryon number, lepton number and spin quantum number, respectively. Then, for SM particles R is equal to 1, and their SUSY partners have R equal to -1. Experimentally, this parity can be illustrated by considering the high stability of protons, which have life-time greater than 10^{32} years as opposed to decay into neutral pion and positron [51]. The R-parity conservation also leads to an important consequence of MSSM: the lightest neutral SUSY particle (LSP) will necessarily be stable and therefore can be a suitable candidate for a dark matter.

5 for a complete review of the supersymmetry please see the review from S.Martin in [50]

In the next few subsections the particle content of the SUSY is discussed, as well as the production and decay of the SUSY particles, and the way to search them at LHC will be addressed.

2.4.1 *Particle content*

Particles in a SUSY theory are represented by the supermultiplets. These multiplets contain both fermion and boson states. Spin-statistics theorem [8] tells us that number of fermionic and bosonic degrees of freedom must be equal. In the simplest case, a supermultiplet then consists of a two-component Weyl fermion[6] and a complex scalar field, the sfermion[7]. It is called a chiral or matter supermultiplet. The ordinary SM fermions fit into these chiral multiplets, even though left- and right-handed fermions need to settle in separate supermultiplets.

A second possibility is to include a spin-1 vector boson. The resulting supermultiplet is called a gauge or vector supermultiplet and consists of a massless spin-1 boson and a massless spin-$\frac{1}{2}$ Weyl fermion. The theory must remain renormalizable, therefore the bosons can only attain a mass through spontaneous symmetry breaking mechanism. Given that gauge bosons transform under the adjoint representation of the gauge group, so must their superpartners, the so-called gauginos. This means that left-handed and right-handed gauginos will transform in the same way under the gauge group as opposed to the fermions we know from the SM.

A first step in construction of a MSSM, is an inclusion of all the SM particles in the supermultiplets. SM fermions – the quarks and leptons – have to be members of chiral supermultiplets, because left- and right-handed fermions transform differently under the electroweak gauge symmetry. Their spin-0 superpartners are called squarks and sleptons. Symbolically, superpartners are denoted with a tilde sign above the usual SM symbol. The chiral supermultiplets are present in the MSSM are summarized in Table 2.1. The representation of the SM gauge group under which the supermultiplets transform is given in the last column. All chiral supermultiplets are defined in terms of left-handed Weyl spinors, which is why the conjugates of the right-handed quarks and leptons appear in the table.

Names		spin 0	spin 1/2	$SU(3)_C$, $SU(2)_L$, $U(1)_Y$
squarks, quarks	Q	$(\widetilde{u}_L \quad \widetilde{d}_L)$	$(u_L \quad d_L)$	$(\mathbf{3}, \mathbf{2}, \frac{1}{6})$
(×3 families)	$\overline{u}$	$\widetilde{u}^*_R$	$u^\dagger_R$	$(\overline{\mathbf{3}}, \mathbf{1}, -\frac{2}{3})$
	$\overline{d}$	$\widetilde{d}^*_R$	$d^\dagger_R$	$(\overline{\mathbf{3}}, \mathbf{1}, \frac{1}{3})$
sleptons, leptons	L	$(\widetilde{\nu} \quad \widetilde{e}_L)$	$(\nu \quad e_L)$	$(\mathbf{1}, \mathbf{2}, -\frac{1}{2})$
(×3 families)	$\overline{e}$	$\widetilde{e}^*_R$	$e^\dagger_R$	$(\mathbf{1}, \mathbf{1}, 1)$
Higgs, higgsinos	H_u	$(H^+_u \quad H^0_u)$	$(\widetilde{H}^+_u \quad \widetilde{H}^0_u)$	$(\mathbf{1}, \mathbf{2}, +\frac{1}{2})$
	H_d	$(H^0_d \quad H^-_d)$	$(\widetilde{H}^0_d \quad \widetilde{H}^-_d)$	$(\mathbf{1}, \mathbf{2}, -\frac{1}{2})$

Table 2.1: Chiral supermultiplets in the MSSM. The spin-0 fields are complex scalars, and the spin-$\frac{1}{2}$ fields are left-handed two-component Weyl fermions. The table is taken from [50].

6 the particle described by a relativistic wave equation for massless spin-$\frac{1}{2}$ particles
7 "s" here stands for scalar

The vector bosons of the SM will have to reside in gauge supermultiplets together with a Majorana fermion[8] field. An overview of the gauge superfields present in the MSSM is given in Table 2.2. After electroweak symmetry breaking, the neutral gauge bosons W^3 and B mix to form the mass eigenstates for Z and γ bosons. The corresponding gaugino mixed states are called the zino ($\widetilde{Z}$) and photino ($\widetilde{\gamma}$).

Names	spin 1/2	spin 1	$SU(3)_C$, $SU(2)_L$, $U(1)_Y$
gluino, gluon	$\widetilde{g}$	g	(**8**, **1** , 0)
winos, W bosons	$\widetilde{W}^{\pm}$ $\widetilde{W}^0$	$W^{\pm}$ W^0	(**1**, **3** , 0)
bino, B boson	$\widetilde{B}^0$	B^0	(**1**, **1** , 0)

Table 2.2: Gauge supermultiplets in the MSSM. The table is taken from [50].

In the MSSM, the superpartners listed in Tables 2.1 and 2.2 are not necessarily the mass eigenstates of the theory. After electroweak symmetry breaking, the remaining symmetry of the theory is the $\mathrm{SU}(3)_c \times \mathrm{U}(1)_{EM}$ symmetry. This means that particles that correspond to the same $\mathrm{SU}(3)_c \times \mathrm{U}(1)_{EM}$ gauge eigenstates can mix with each other. The mass eigenstates can be determined by diagonalizing the mass matrices, which is constructed by gathering all terms in the Lagrangian that have a quadratic component of the fields.

The charged electroweak gauginos, $\widetilde{W}^1$ and $\widetilde{W}^2$ (the winos), and the charged higgsinos, $\widetilde{H}_u^+$ and $\widetilde{H}_d^-$, mix to form the so-called charginos denoted by $\widetilde{\chi}_1^{\pm}$ and $\widetilde{\chi}_2^{\pm}$. The neutral wino ($\widetilde{W}^3$) and bino ($\widetilde{B}$) and the higgsinos ($\widetilde{H}_u^0$ and $\widetilde{H}_d^0$) form neutralinos, denoted as $\widetilde{\chi}_1^0$, $\widetilde{\chi}_2^0$, $\widetilde{\chi}_3^0$ and $\widetilde{\chi}_4^0$. The lightest among the neutralinos - $\widetilde{\chi}_1^0$ - is considered as the LSP.

The gluino, which is a color octet, cannot mix with any other particle. In this respect it is unique among all MSSM particles. The mass eigenstate is thus the same as the gauge eigenstate. The squarks and sleptons are scalars with the same quantum numbers, and so they can mix with each other. The mass eigenstates of the squarks and sleptons should therefore be determined by diagonalizing three 6×6 mass matrices – one for up-type squarks, one for down-type squarks and one for charged sleptons – and one 3×3 mass matrix – for the sneutrinos. Because of the dependence on the fermion masses, the mixing is especially important for the top squark sector. The mass eigenstates for the top squark are usually denoted by $\widetilde{\mathrm{t}}_1$ and $\widetilde{\mathrm{t}}_2$. The $\widetilde{\mathrm{t}}_1$ can become substantially lighter than all other squarks because of the large mixing that can be present. Therefore, it is very well possible that the $\widetilde{\mathrm{t}}_1$ is the only accessible squark at the LHC. The squarks of the first and second generation are usually grouped together, and assumed to be mass-degenerate.

2.4.2 *Particle production and decay*

When we assume that R-parity is conserved, the production of sparticles at hadron colliders such as the LHC can only occur in pairs. The production of squarks and gluinos will proceed predominantly via the strong interaction while neutralinos, charginos and sleptons can only be produced via the electroweak interaction. An overview of the cross sections for the production of various sparticles is shown in Fig. 2.7.

8 Majorana particle is a fermion that is its own antiparticle

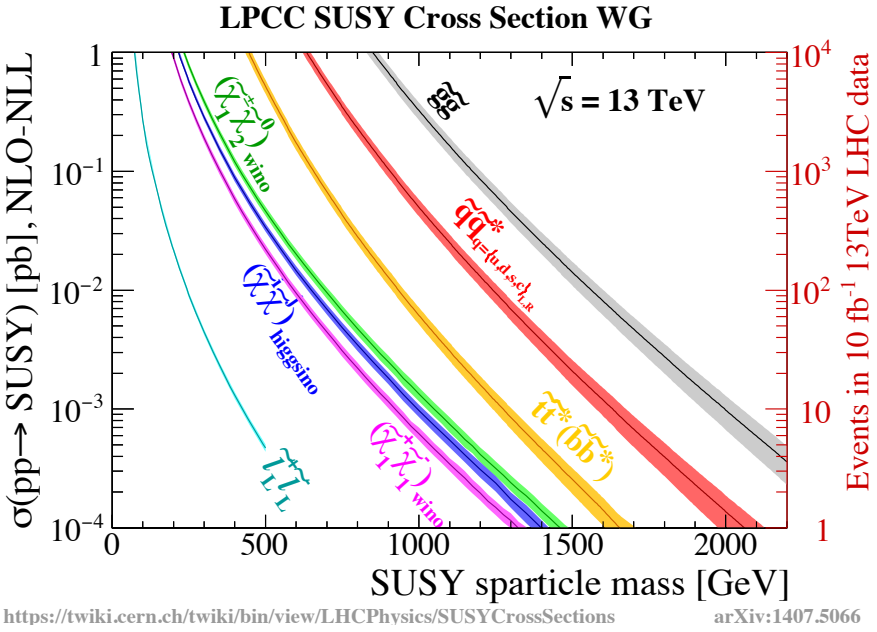


Figure 2.7: Cross sections for the production of various supersymmetric particles at a center-of-mass energy of 13 TeV. The expected number of produced events for a dataset collected in 10 fb^{-1} at 13 TeV center-of-mass energy are also shown. The figure is taken from [52].

The gluino pair production is the dominant one, followed by the squark production. The electroweakinos have a much lower cross section. At the LHC, the production of squarks and gluinos is dominated the gluon-gluon and gluon-quark fusion. Once produced, the SUSY particles decay in a long decay cascade in which quarks and leptons can be produced. The full decay chain ends with the production of at least two LSPs to fulfill the requirement for the R-parity conservation. The LSP escapes from the detector without being detected, which will result in an imbalance of transverse momentum - one of the most powerful discriminating variables in the search for SUSY signals.

In general, SUSY signatures can contain any number of jets, leptons, and gauge bosons, depending on the details of the mass spectrum. The SUSY search program in the CMS experiment therefore covers a wide variety of final state topologies. Each analysis will be more or less sensitive to certain classes of SUSY parameter points. Therefore, the main challenge for the SUSY searches - importance to search everywhere, in all possible ways.

The decay possibilities for SUSY particles depend strongly on the precise SUSY mass spectrum. This spectrum is very model-dependent, and is largely related to how SUSY has been broken, and which relationships between masses and couplings are assumed to be realized. If kinematically allowed, two-body decays will always dominate. In case the mass spectrum is more compressed, and the direct two-body decay is forbidden, the decay will proceed via virtual SUSY particles leading to three- and four-body decays.

Gluinos can only decay through a squark, which can be on-shell or not. As discussed above, the top and bottom squarks can be much lighter than the other squarks, which means that $\widetilde{g} \to \mathrm{t}\widetilde{\mathrm{t}}_1$ and $\widetilde{g} \to \mathrm{b}\widetilde{\mathrm{b}}_1$ could be the only available two-body decays. In this case, they will dominate and b tagged jets will appear in the final states. In case all squarks are heavier than the gluino, it will decay via a three body decay to two quarks and a neutralino or chargino. For squarks, the decay to quark and gluino will be dominant if the gluino is light enough. In case this decay is not accessible, the decay can proceed

via the electroweak interaction to a quark and neutralino or chargino. The neutralinos and charginos are an admixture of the electroweak gauginos and therefore inherit their couplings. If neutralino or chargino appears in the decay chain it will automatically decay to a lighter neutralino and a W or Z boson.

2.4.3 *Simplified model spectra*

Many extensions of the SM are possible, and many of the models have numerous free parameters. The MSSM is not an exception. Without additional simplifications there are 104 extra free parameters; with a minimal set of assumptions, there are still about 20 parameters left. This extremely large parameter space is not very practical to do an analysis. Furthermore, similar experimental signatures can be produced in different ways, and by different models. To address these issues, a simplified approach was developed, using the so-called simplified model spectra (SMS) [53–55]. These SMS are based primarily on the experimental topologies, how many jets or leptons are produced, whether there are jets originating from b quarks, and so on. Simplified models can only contain a limited number of particles and interactions. This will result in a very minimal set of parameters for which no assumptions on relations between parameters are made. In short, it could be said that these SMS models are more suitable to optimize and interpret a new physics search compared to a full-developed NP model.

Part of this thesis is dedicated to the searches for strong SUSY production. Four SMS SUSY models featuring gluino pair or 3rd generation squark pair production and assumed R parity conservation are used to perform a search in a multi-lepton final state. These studies are discussed in Chapter 6.

3

THE CMS EXPERIMENT AT THE LHC

In order to study the physics described in Chapter 2, we need a particle accelerator which provides us with high-energy proton-proton collisions, in which the physics processes of interest are produced. The proton-proton collision events provided by the Large Hadron Collider (LHC) [56], the most powerful collider built by human, are used in the study presented in this thesis. The LHC machine details will be described in the first part of this chapter. Smashing particles in the laboratory has its benefit: we can control what happens at the collision point of the beams by reconstructing the collision products. For this purpose, we use the Compact Muon Solenoid (CMS) [57] detector, which is described in the second part of this chapter.

Colliders may either be ring accelerators or linear accelerators, and may collide a single beam of particles against a stationary target or two beams head-on. When two beams collide, the center-of-mass energy of the collision is the sum of their energies. A beam of the same energy that hits a fixed target would produce a collision of much less energy which is proportional to square root of the energy beam in the center-of-mass frame. A linear particle accelerator is a type of particle accelerator that accelerates charged subatomic particles by putting them to a series of oscillating electric potentials along a linear beam line. The most famous one in this category is the Stanford Linear Collider (SLC) accelerator, that collided electrons and positrons with the center-of-mass energy close to the Z boson mass, thereby investigating its properties. In a circular accelerator, particles are accelerated in a circle until they reach sufficient energy. The advantage of circular accelerators over linear accelerators is that the ring topology allows continuous acceleration, as the particle can transit indefinitely. Besides the LHC, other remarkable circular accelerators were the Tevatron, which discovered the top quark in 1995, and KEKB, which until 2018 had the world's highest luminosity record. Besides these colliders, another remarkable and unique accelerator was hosted by DESY in Hamburg. HERA was an electron-proton accelerator, which had the investigation of the proton structure as its main task.

To accelerate a particle, it has to have an electromagnetic charge, which can be used to accelerate and focus the beam of the particles. Another important feature is that the particle should be stable enough so it will not decay during the acceleration time. Among all particles, the best candidates for the accelerators are electrons, protons, and heavy nuclei. The proton has its internal structure, it consists of so-called partons, which are carrying part of the total energy of the proton. Because the initial momenta of the partons are not fixed values, proton-proton interactions can be used to probe a broad range of energies. Thus proton-proton colliders can be used as discovery machines. On the contrary, various experiments show that there is no structure for the electron. This means we know the initial energy of interaction in electron colliders with extreme precision.

Of course, nothing comes without a price: in circular colliders, particles are bend by magnets to stay inside the collider trajectory. When a particle changes trajectory it loses energy due to the so-called synchrotron radiation or bremsstrahlung. The intensity of this radiation is inversely proportional to the fourth power of the mass of the particles, which means electrons lose significantly larger amount of energy than protons of the same energy. For example, the last largest electron-positron collider - LEP - had severe constraints on the possible maximum energy of the electrons. At the end of the LEP's run, each revolution caused electrons to lose approximately 3% of their energy [58]. This constraint kept LEP from discovering the BEH boson, which was found by the ATLAS and CMS Collaborations in 2012.

3.1 LHC

To obtain protons for further acceleration, hydrogen gas is injected into a metal cylinder to break down the gas into protons and electrons. The protons are then sent to the linear accelerator LINAC 2, where they are accelerated to 50 MeV and then injected into the Proton Synchrotron Booster, which accelerates them up to about 1.4 GeV. At the next step, protons are transferred to the Proton Synchrotron, which increases their energy up to 25 GeV and afterward the Super Proton Synchrotron ring brings the protons energy up to 450 GeV. The LHC is the last ring in a complex chain of accelerators, where proton bunches reach the energy of 6.5 TeV and cause them to collide head-on, creating center-of-mass energies of 13 TeV. The entire accelerator complex is shown in Fig. 3.1.

In the accelerator, particles circulate in a vacuum and are steered using electromagnetic field: dipole magnets keep particles at their nearly circular orbits, quadrupole magnets focus the beam, and accelerating cavities are electromagnetic resonators that accelerate particles and keep them at the fixed energy by compensating for energy losses. Protons in phase with the electric field are accelerated, while the protons out of phase with the field are decelerated; this mechanism allows to group protons into several bunches. The LHC uses eight cavities per beam, each capable of delivering 5 MV/m accelerating field at 40 MHz, as a consequence these bunches are separated by a 25 ns spacing. The 15-meters long dipole magnets are based on superconducting technology and designed to generate a high field of 8.3 T. In order to achieve the strength of the magnetic field the niobium-titanium (NbTi) cables are used, which are operated at a temperature of 1.9 K. In LHC, there are 1232 dipole and 400 quadrapole magnets, which under nominal operating conditions transport 2×2808 bunches in the opposite directions, each containing 10^{11} protons.

3.1.1 *Luminosity*

The beams are colliding at several interaction points along the LHC; apart from CMS, these are ATLAS, ALICE and LHCb experiments. The number of interactions that effectively takes place at these collision points is determined by the instantaneous luminosity, L, which is defined as the ratio of the number of events detected in a certain time period to the interaction cross section:

$$L = \frac{1}{\sigma}\frac{dN}{dt} \tag{3.1}$$

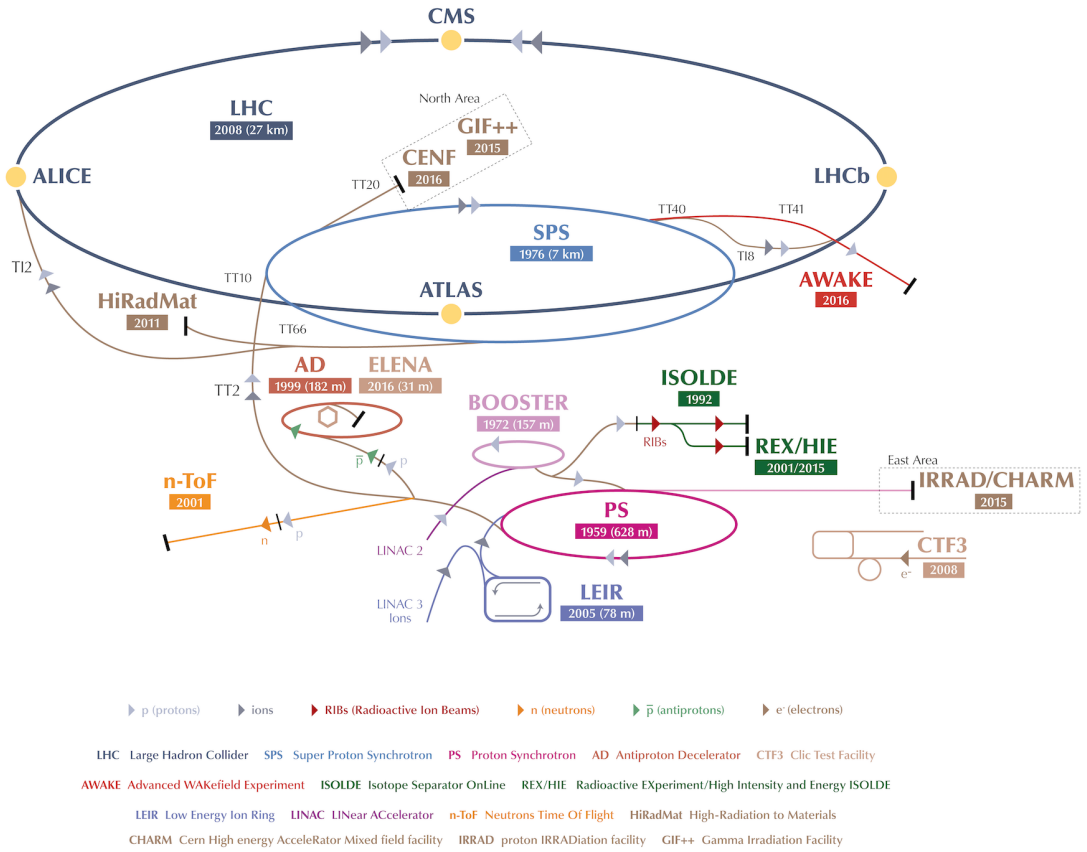


Figure 3.1: Schematic representation of the CERN accelerator complex and experiments(taken from [59]).

Integrating this over time gives a related quantity - integrated luminosity:

$$\mathcal{L} = \int L dt \tag{3.2}$$

These parameters are useful to characterize the performance of an accelerator. Once the integrated luminosity is known, the expected number of events of a given type, having certain cross section, σ, is calculated as:

$$N = \sigma \mathcal{L} \tag{3.3}$$

If two bunches containing n_1 and n_2 protons collide head-on with frequency, f, an instantaneous luminosity can be expressed as:

$$L = f \frac{n_1 n_2}{4\pi \sigma_x \sigma_y}, \tag{3.4}$$

where σ_x and σ_y characterize transverse beam sizes in the horizontal and vertical directions.

In the second run period (Run II) the LHC reached its peak instantaneous luminosity of about $2.1 \times 10^{34} cm^{-2} s^{-1}$, which is higher than any previous machine operating at

the energy frontier. The LEP, for instance, had an instantaneous luminosity of about $2 \times 10^{31} cm^{-2} s^{-1}$ [60] while the Tevatron reached $4 \times 10^{32} cm^{-2} s^{-1}$ colliding protons and antiprotons. In 2018 the LHC had overcome the record of the highest instantaneous luminosity out of any accelerator; the previous one was held by the KEKB for almost 10 years [61].

As it can be seen from Eq. 3.2 the instantaneous luminosity can be described as the "data collection speed". As an example, the ZEUS detector at HERA accelerator was operating from 1992 to 2007. During these 15 years, it collected 0.5 fb^{-1} of data [62], which is an amount of the data collected by the CMS detector each operational day in 2018 [63].

3.1.2 *Pile-up*

Every time two proton bunches collide, multiple proton-proton collision takes place. This effect is known as pile-up (PU). The CMS detector has to register decay products of unstable particles and reconstruct backward which particles correspond to different primary vertexes. The CMS detector was designed to operate at an average number of 25 interactions per crossing and with instantaneous luminosity of $\approx 1 \times 10^{34} cm^{-2} s^{-1}$. Currently, this number has been already exceeded by factor 2. Such an instantaneous luminosity causes an intense particle flux through the CMS sub-detectors and might cause severe radiation damage in the detector elements. The PU distribution in the CMS detector is shown in Fig. 3.2 (left) for the collisions which happened in 2018. As an example, Fig 3.2 (right) shows a recorded event that contained 78 simultaneous proton-proton collisions in a single bunch crossing.

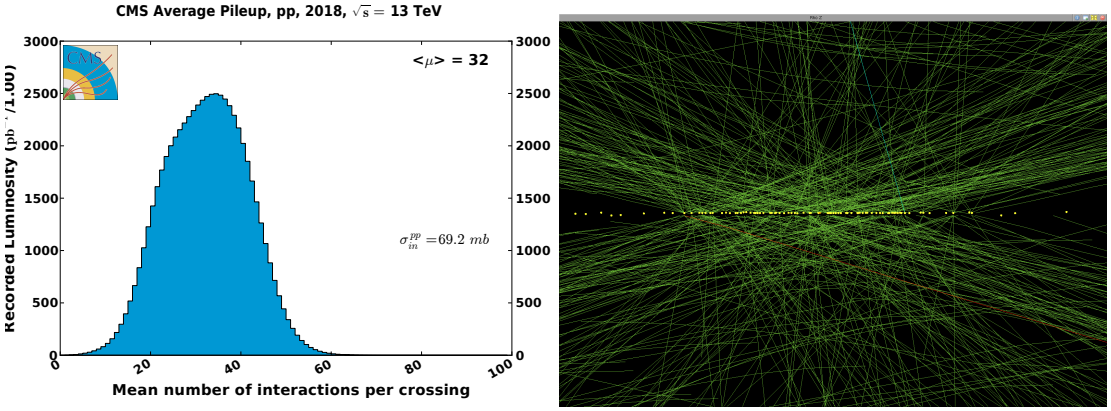


Figure 3.2: Distribution of the number of proton-proton collisions per bunch crossing at the CMS interaction point during the 13 TeV run (left) [63], and an event display of an 8 TeV event, where 78 distinct vertexes were reconstructed. Data recorded on the 21st of September 2012 (right) [64].

3.2 THE CMS DETECTOR

CMS together with ATLAS are two general purpose detectors specifically designed to study the BEH boson and BSM physics. The CMS detector has a cylindrical shape, around LHC's beamline, consisting of a central barrel and two endcaps. CMS is 21.6 m long, with a diameter of 15 m. The central element of the CMS apparatus is a superconducting solenoid, providing a magnetic field of 3.8 T. CMS comprises of various sub-detectors

and technologies, which are shortly described in this section. A silicon pixel and strip tracker, a lead tungstate ($PbWO_4$) crystal electromagnetic calorimeter (ECAL), and a brass and scintillator hadron calorimeter (HCAL) are all placed within the CMS solenoid. Muons are detected in gas-ionization chambers embedded in the steel flux-return yoke outside the solenoid. A graphical view of CMS with its sub-detectors is given in Fig. 3.3.

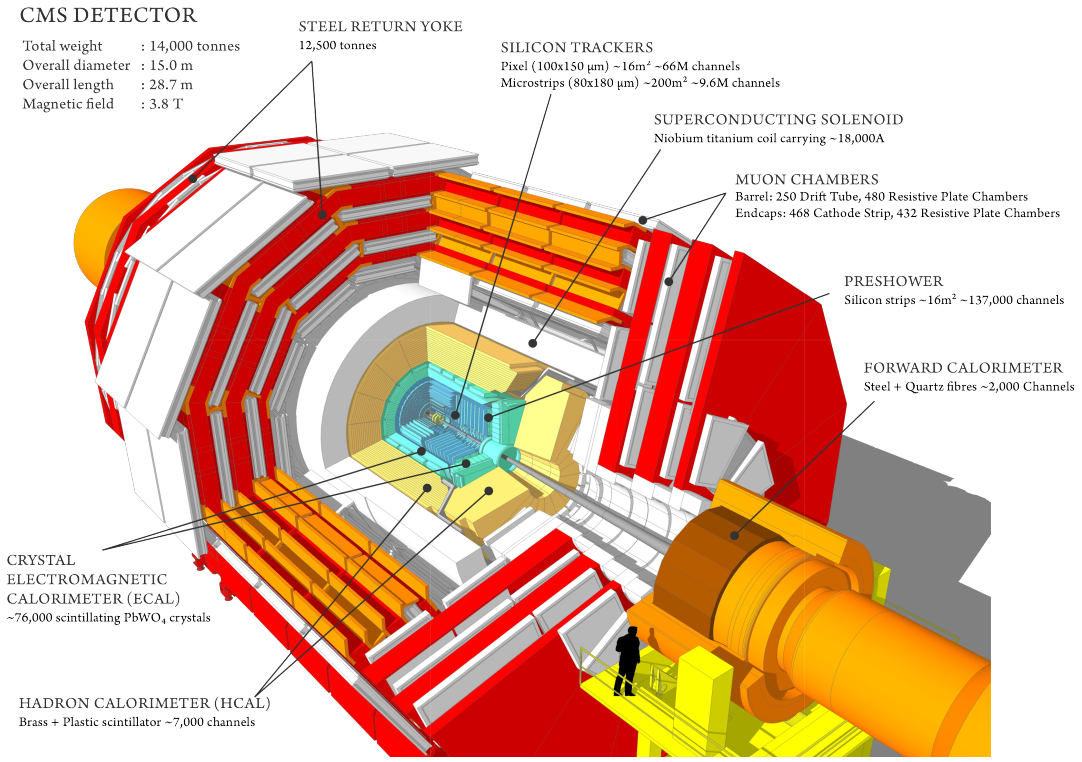


Figure 3.3: Schematic view of the CMS detector. The LHC beams travel in opposite directions along the central axis of the CMS cylinder colliding in the middle of the CMS detector. The figure is taken from [65].

3.2.1 *CMS coordinate system*

CMS uses a right handed Cartesian coordinate system where the origin is at the nominal collision point inside CMS. The z-axis points along the tangent to the counter-clockwise rotating proton beam, the x-axis is pointing to the center of the LHC and y-axis pointing up perpendicular to the LHC plane. These coordinate positions are indicated in Fig. 3.4.

The polar angle, θ, is measured from the positive z-axis and the azimuthal angle, ϕ, is measured from the positive x-axis in the x-y plane. One of the important quantities - pseudorapidity, η, is defined as

$$\eta = -\log\left(\tan\left(\frac{\theta}{2}\right)\right)$$

and often used instead of the polar angle. The reason for this is that the difference in rapidity between two particles, $\Delta\eta$, is invariant under longitudinal Lorentz boosts

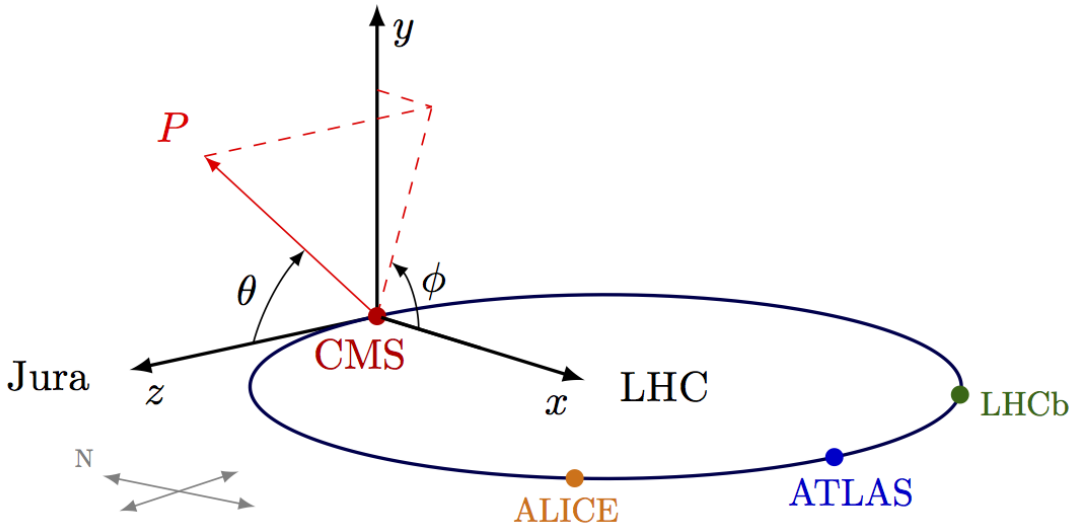


Figure 3.4: Coordinate system used by the CMS experiment at the LHC. Here Jura indicates a direction of a sub-alpine mountain range located north of the Western Alps. The Figure is made with a code taken from [66].

along the beam axis. The angular separation between two detected particles is usually expressed in terms of the longitudinal Lorentz boost invariant quantity ΔR, defined as:

$$\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2}. \tag{3.5}$$

Here $\Delta\phi$ is difference in ϕ coordinate between two particles and this quantity is essential to be invariant under longitudinal boosts since its defined in the plane orthogonal to the beam direction [57]. As already mentioned, CMS detector has a cylindrical shape, so it's essential to use one of the variables from cylindrical coordinate system. The axial distance or radial distance, r, is picked as the Euclidean distance in the x-y plane.

The momentum of a particle can be decomposed into longitudinal, parallel to the beam axis, and a transverse, perpendicular to the beam axis in the x-y plane $\left(p_{\mathrm{T}} = \sqrt{p_x^2 + p_y^2}\right)$, components. When two partons collide at CMS, the initial momentum is not known along the beam axis direction (the particle can go undetected inside the beam pipe), while the momentum balance in the transverse plane must remain zero. Using the latter fact, we define the momentum and energy in the transverse plane, p_{T} and E_{T}, respectively. Particles which leave the detector undetected could be revealed by the missing energy p_{T}^{miss} in the transverse plane; this could be for instance neutrinos.

3.2.2 *Detectors within CMS*

In the following, we describe the CMS sub-detectors starting from the closest one to the collision point and then ordered by distance increasing from the interaction point.

TRACKER The tracking detectors are dedicated to provide a precise and efficient measurement of the charged particle momentum and trajectories. In addition, high granularity of the tracking system allows to reconstruct precisely the primary interaction vertex. It is usually taken as the one corresponding to the hardest proton-proton

interaction, and distinguished from those coming from PU vertexes and secondary vertexes from decays of long-lived heavy particles.

The tracker has a total length of 5.8 m and a diameter of 2.5 m and has a coverage in $|\eta|$ up to 2.5. A schematic overview of its geometry is shown in Fig. 3.5. Within a few cm from the interaction point, a silicon pixel detector is installed to cope with the high particle flux. Around the pixel detector, silicon micro-strips are used. When charged particles pass through these pixels and strips, they cause ionization in the silicon, creating numerous electron-hole pairs. When an electric field is applied, these electrons and holes in the silicon drift towards the electrodes, generating fast signal in the detector.

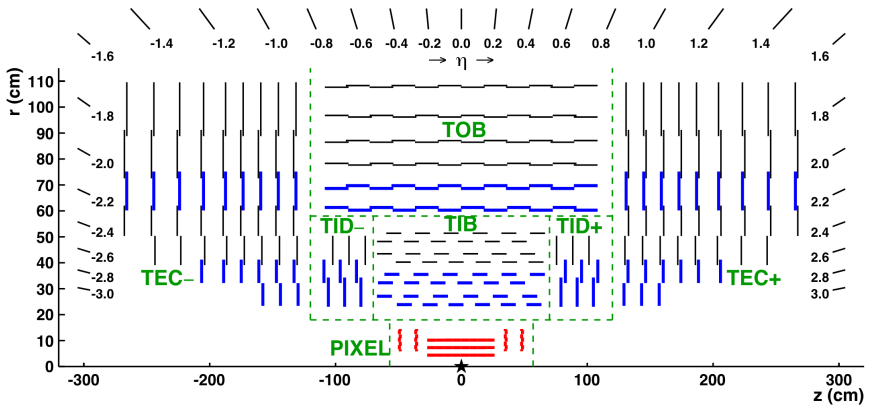


Figure 3.5: Schematic view of the CMS tracker in half of the r-z plane. The center of the tracker, corresponding to the approximate proton-proton collision point, is indicated by a star. Black and blue lines show strip tracker modules. Red lines show the pixel modules. The meaning of all abbreviations is described in the text. The Figure is taken from [67].

The pixel detector consists of three barrel layers at radii between 4.4 and 10.2 cm from the beam axis, and at 34.5 and 46.5 cm from the interaction point in forward regions. The size of the silicon pixels, $100 \times 150\ \mu m^2$ in r-ϕ and z planes, is chosen to achieve an excellent spatial resolution ($10\ \mu m$ in r-ϕ plane and $15\ \mu$ along z-axis) and to keep the detector occupancy at a maximum of 1% and efficiency well above 99%. The pixel detector covers η range up to 2.5. The original pixel detector was designed to withstand instantaneous luminosity of $1 \times 10^{34} cm^{-2} s^{-1}$, while the luminosities of $1.5 \times 10^{34} cm^{-2} s^{-1}$ were already achieved in 2016 and some performance degradation was observed with a time. Therefore, the decision was made to replace the pixel detector during the technical stop in the beginning of 2017. The upgraded pixel detector has one more sensitive layer in both barrel and forward regions, being designed to operate at a particle flux up to $2 \times 10^{34} cm^{-2} s^{-1}$ [68].

The strip tracker is composed of silicon micro-strip detector and is divided in an inner region, with 4 barrel layers and 3 endcap layers, and an outer region, with 6 barrel layers and 9 endcap disks. The modules in the inner rings (denoted as TIB and TID in Fig. 3.5) and first four endcap modules (denoted as TEC) use 320 μm thick silicon sensors, while those in the outer rings (denoted as TOB) and the outer three endcap rings use silicon of 500 μm thickness. Additionally, some of the layers carry a second microstrip detector, tilted with respect to each other, in order to measure the second coordinate, respectively z and r for the barrel and endcap. Within a layer, each module is shifted slightly in z or r with respect to its adjacent modules, allowing for an overlap to avoid gaps in acceptance.

ECAL The CMS ECAL is a compact, hermetic, fine-grain, homogeneous calorimeter made of lead tungstate scintillating crystals [69]. The choice of lead tungstate was primarily motivated by its radiation tolerance, its small radiation length (0.89 cm) and Moliere radius[1], and its fast response (99% of the light is collected in 100 ns, which is compatible with the 25 ns bunch spacing at LHC). The cylindrical barrel consists of 61200 crystals and the ECAL endcaps are made up of 15000 crystals. The crystal length is 23 cm in the barrel and 22 cm in the endcap. When electrons or photons pass through the ECAL, they collide with the nuclei of the crystals and deposit their energy by bremsstrahlung or pair production generating a scintillation light in the crystals. The scintillation light output is rather low, so highly efficient photodetectors are placed at the rear of the crystal to increase the current produced by scintillation light.

The ECAL is composed of a barrel, surrounding the tracker, and two endcap sections. The barrel part of the ECAL covers a range in η up to 1.479 and the endcaps cover the rapidity range $1.479 < |\eta| < 3.0$. The $1.653 < |\eta| < 2.6$ range is also equipped with a preshower detector with high granularity, able to distinguish a neutral pion decaying to photons from actual photons. The preshower has a much finer granularity than the ECAL with detector strips of 2 mm wide, compared to the 3 cm wide ECAL crystals, and can register each of the pion-produced particles as a separate photon. The CMS ECAL layout is shown in Fig. 3.6.

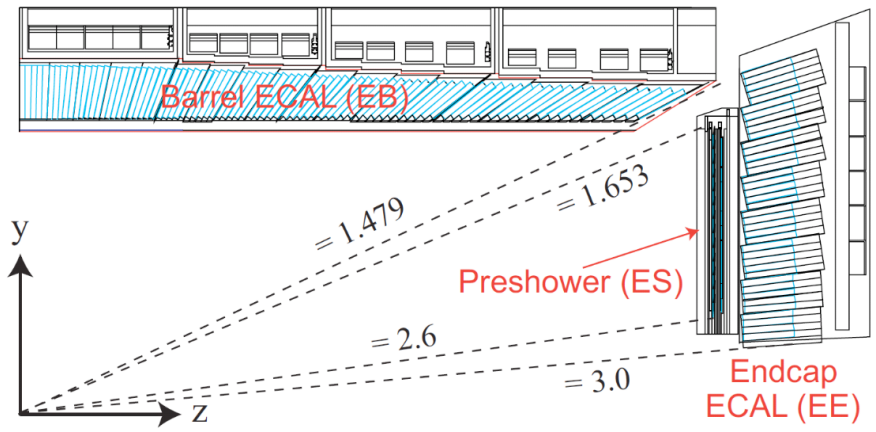


Figure 3.6: Geometric view of one-quarter of the ECAL in the y-z plane. The numbers on the plot reflect the position of the ECALs components as a function of η. The Figure is taken from [70].

The ECAL energy resolution for electrons in barrel region has been measured in the beam tests and can be approximated as:

$$\frac{\sigma_E}{E} = \frac{2.8\%}{\sqrt{E(\text{GeV})}} \oplus \frac{12\%}{E(\text{GeV})} \oplus 0.3\%, \tag{3.6}$$

where the first is a stochastic term responsible for intrinsic fluctuations from shower development, second is the noise term due to the electronic noise of the readout chain and third is the constant term that is usually caused by instrumental effects [71].

HCAL The HCAL's purpose is to measure the energy of both neutral- and charged hadrons. When hadrons enter a medium they initiate a hadronic cascade, known as a

1 the radius of a cylinder containing 90% of the electromagnetic cascade's energy deposition

shower, through strong interactions with the nuclei. These showers develop relatively slow compared to electromagnetic cascades, and even though the hadrons already lose some energy in the ECAL, they will deposit most of it in the HCAL. As opposed to the ECAL, which is a homogeneous detector, the HCAL is a sampling calorimeter which consists of alternating layers of high density absorber material and layer of plastic scintillators converting the absorbed energy into the light pulse.

The detector is designed to be as hermetic as possible and covers a range in η up to 5.0. This is to ensure there are no particles escaping from the detection which is essential for the calculation of p_{T}^{miss}. The HCAL consists of four different detectors. Inside the magnetic coil are the barrel detector (HB), covering a range in η up to 1.3, and the endcap detector (HE) which covers the range $1.3 < \eta < 3.0$. The outer detector (HO) is an additional layer of scintillators placed outside of the magnet, in order to collect the energy from hadron showers penetrating through the barrel detector. Finally, coverage between $2.9 < |\eta| < 5.0$ is guaranteed by the HF detector, which detects particles through the emission of Cherenkov light in the absorber crystals. The geometry and location of the HCAL sub-detectors is shown in Fig. 3.7.

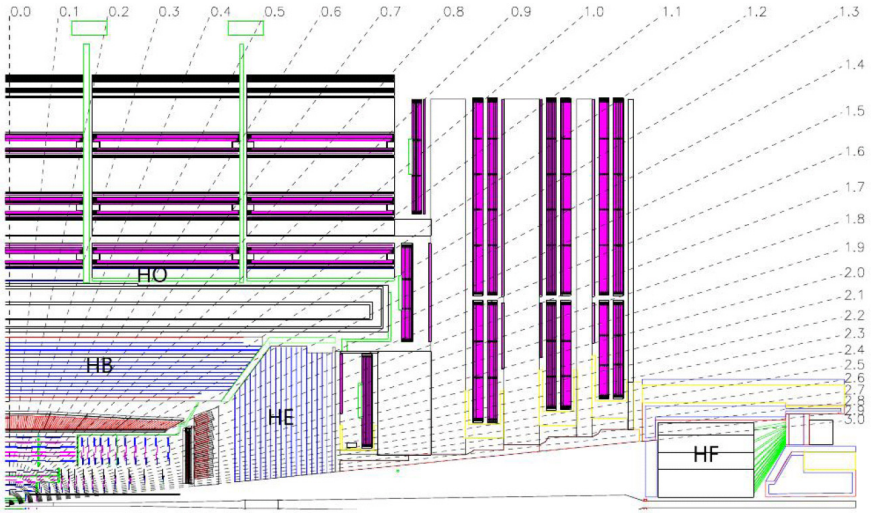


Figure 3.7: Longitudinal view of the CMS HCAL detector showing the locations of the barrel (HB), endcap (HE), forward (HF) and outer (HO) detectors. The Figure is taken from [72].

The energy resolution for the reconstructed jet in the HCAL at energies around 30 GeV is about 30%, and it improves for increasing energy to less than 10%, in all regions of the detector [57].

MAGNET AND RETURN YOKE The CMS detector uses a solenoid magnet which is the core of the detector, and is also reflected in the name of the experiment. A good momentum resolution of the high momentum charged particles is achieved by applying a strong magnetic field and measuring the curvature of the charged particle track. The CMS magnet is 6 m in diameter and 12.5 m in length and is made out of NbTi. The temperature of 4 K provided by the cooling system ensures the magnet to stay in a "superconducting" regime, i.e. electricity flows without resistance inside the magnet, and creates a powerful 3.8 T uniform magnetic field. The magnet return yoke of the CMS detector, confines the magnetic field and stops all remaining particles except for muons and neutrinos.

MUON SYSTEM Muons can penetrate several meters of iron, losing their energy only via ionization, and contrary to other particles they are not absorbed by any of CMS sub-detectors. Therefore, chambers to detect muons are placed at the very end of the CMS detector. The muon system is composed of three types of gaseous detectors placed between the return yoke layers.

When a muon traverses a gas chamber, it ionizes the gas creating electron-ion pairs. Applied voltage between electrodes creates an electric field, where electrons drift to the positively charged anode whereas ions slowly move to the negatively charged cathode. Close to anode surface, electron energy becomes sufficient to ionize other atoms in the gas. As a result the avalanche inside the gas volume is created, which amplifies the output signal.

In the barrel region ($|\eta| < 0.9$), where the muon rate is low and the magnetic field uniform, drift tubes (DTs) are used. The DTs are installed in the 4 detector stations alternated with the segmented return yoke. Each station is arranged in 2 or 3 so-called "superlayers", each consisting of 4 layers of DTs. The "superlayers" are orthogonal to each other, in which each "superlayer" focuses on the measurement in the direction of ϕ or z.

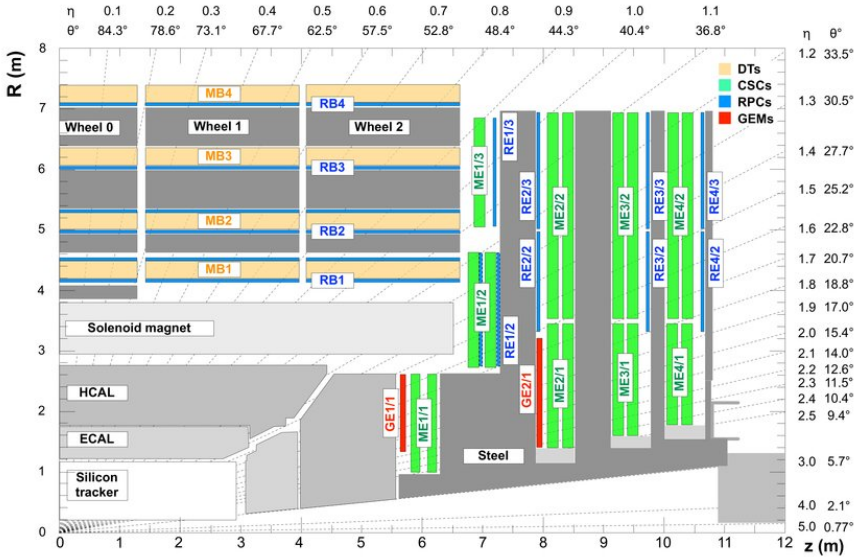


Figure 3.8: The quadrant of the CMS muon system in r-z plane. The steel yoke is represented by darkly shaded blocks between the muon chambers. The DTs are shown in light orange, CSCs in green, RPCs in blue and GEMs in red. The "B" letter in the name indicates the chamber belongs to the barrel and "E" represents the endcap part. The Figure is taken from [73].

In the endcap region ($0.9 < |\eta| < 2.4$), where particle rates are higher and the magnetic field is non-uniform and large, cathode strip chambers (CSCs) are used. CSCs have a trapezoidal form and consist of a cathode plane divided into strips and multiple anode wires orthogonal to the cathode strips. The cathode strips measure the ϕ coordinate, while the anode wires are optimized for bunch crossing identification and measurement in the η direction.

Resistive plate chambers (RPCs) are used in both barrel and endcap regions up to $|\eta| < 2.1$. RPCs consist of two gas gaps, one serves as anode and another as a cathode. Due to the electric field in the gap, an electron avalanche is created when an ionizing particle crosses the detector. The movement of the charges induces a signal on the read-

out strips. RPCs have a coarser position resolution compared with the other two muon detectors, but their fast response allows to use them as a trigger for muon identification.

A high-luminosity upgrade of the LHC is foreseen during a third long shutdown, which starts in 2023, to further increase the instantaneous luminosity to $5 \times 10^{34} cm^{-2} s^{-1}$. To cope with this unprecedented luminosity the CMS muon system must be able to withstand a physics program that maintains sensitivity for electroweak scale physics and for TeV scale searches. One of the future improvements in the muon system is the installation of additional layers of muon detectors which are based on gas electron multiplier (GEM) technology in the first endcap muon station in order to maintain and improve the forward muon triggering and reconstruction in the region $1.6 < |\eta| < 2.2$. The proposed GEM detectors satisfy all performance requirements and are able to operate at particle fluxes far above those expected in the forward region under high luminosity LHC conditions [74].

3.3 TRIGGER SYSTEM AND DATA ACQUISITION

When CMS is running at design luminosity, about one billion proton-proton interactions take place every second inside the detector. Such amount of data could not be processed and stored for all collisions. The vast majority of interactions are low-energy soft collisions, while interesting, energetic head-on collisions happen much less often. Therefore, CMS designed a powerful multilayer trigger system to select interesting events.

The trigger system consists of two levels designed to select events corresponding to rare physics processes, e.g. BEH and BSM physics, from billions of proton-proton collisions. The first level of the trigger is implemented on a hardware level, and selects events containing detector signals consistent with an electron, photon, muon, τ lepton, jet, or p_{T}^{miss} signatures. It reduces the initial event rate to an output rate of 100 kHz and has only 3.2 μs to accept or reject an event. As it is impossible to readout full tracker in this short time, the first level trigger relies solely on the information from the muon system and the calorimeters.

If an event is accepted by the first level trigger, data from all sub-detectors are analyzed at the High Level Trigger (HLT), where the event rate gets reduced to 1-2 kHz. The HLT trigger algorithms are built up of a series of filters to make a decision on an event. To pass the trigger, an event needs to pass every filter in the corresponding trigger path. Events passing the HLT are recorded permanently for further physics analysis [75].

3.4 GRID COMPUTING

CMS produces about five petabytes of data per year running at peak performance, even though this amount is already significantly decreased by the trigger system. To handle this amount of data, a distributed computing and data storage infrastructure called the Worldwide LHC Computing Grid was developed [76]. The grid gets together tens of thousands of standard computers worldwide, giving access to data and simulation samples and analysis of the data to thousands of scientists all over the world.

4

EVENT GENERATION

The proton-proton collisions at the LHC can be categorized into three types of collisions: elastic scattering, diffractive and non-diffractive inelastic processes. In this chapter, we first describe non-diffractive collision processes relevant for this PhD thesis, and later give an overview of particle physics generators.

4.1 HADRON COLLISIONS

One type of interaction between protons is an elastic scattering. In this process, the kinetic energy of a particle is conserved in the center-of-mass frame and only momentum is exchanged between the protons. As the two protons stay intact, elastic scatterings do not involve an exchange of quantum numbers between the protons.

Diffractive processes are similar to elastic scatterings as they do not exchange quantum numbers between the protons, but the transferred momentum causes the single or double diffractive dissociation of protons. There is a third topology, called central diffraction, where an excited system between the protons is created.

In non-diffractive inelastic processes, none of the protons survive in the collision, and the interaction takes place actually between its constituents, so-called partons. Most of these interactions occur with low momentum transfer. However, sometimes partons could carry a large fraction of proton momentum resulting in a hard scattering event. This type of events can lead to the production of heavy SM and BSM particles.

According to asymptotic freedom in QCD, hadrons interact weakly at high energies, corresponding to a smaller coupling constant, α_s, so that the constituents of the hadron can be considered as free particles. Whereas, at low energies the interaction becomes strong, leading to the confinement of quarks and gluons within composite hadrons [77, 78]. The high energy interactions can be calculated perturbatively while for low energies it is not possible due to large α_s. The probability, i.e. cross section, of a hard inelastic scattering process AB $\rightarrow$ X can be estimated using the factorization scheme, in which the perturbative description of a hard scattering parton-parton interaction is separated from the non-perturbative contribution to the process:

$$\sigma = \sum_{a,b} \int_0^1 dx_a dx_b \int f_a^A(x_a, \mu_F) f_b^B(x_b, \mu_F) d\hat{\sigma}_{ab\rightarrow c}(\hat{s}, \mu_F, \mu_R). \tag{4.1}$$

Here f_a^A is the parton distribution function (PDF), describing the probability to find a parton of type a (e.g. quark, anti-quark, gluon) carrying a momentum fraction of the proton, x, in a hadron A. The PDFs can not be obtained via perturbative QCD calculations. However, they are independent of the process under consideration and hence universal to all high energy physics experiments. Therefore, the f_a^A are systematized by several experimental collaborations, based on data from previous and current experiments.

The PDFs are also dependent on the momentum scale at which the hadron is probed: the value of the coupling constant, α_s, affects gluon emission and gluon splitting processes, which affect the quark and gluon parton distributions. A factorization scale, μ_F, is introduced at which the f_a^A are extracted. Emissions with transverse momenta below μ_F are absorbed into the f_a^A, while emissions at higher transverse momenta are supposed to be calculated perturbatively. Fortunately, it is possible to describe the dependence of these splitting processes as a function of μ_F, resulting in the DGLAP evolution equations [79–81] which enable the extrapolation of the parton distribution functions from one scale to another.

The hard interaction differential cross section for the partons a and b to some final state c is denoted by $d\hat{\sigma}_{ab\to c}$. This term contains only hard emissions above the factorization scale μ_F and can be calculated in terms of perturbative theory. The term μ_R is the renormalization scale, which is used to make sense of infinite integrals in perturbation theory and $\hat{s}$ is the parton center-of-mass energy. The differential cross section term can be reduced to the formulation of the Matrix Element (ME), which is a sum over the Feynman diagrams [19] contributing to the scattering amplitude. The ME enters the cross section formula as its square. As a result, the total cross section of different processes leading to the same final state is not simply the sum of the cross sections of the individual contributing processes, but also positive or negative interference effects are taken into account.

The same final state X can be created by different combinations of incoming parton types a and b. Therefore, one needs to sum over all parton types and integrate over their possible momentum fractions x_a and x_b in order to include all possible contributions. A graphical representation of what described earlier is shown in Fig. 4.1.

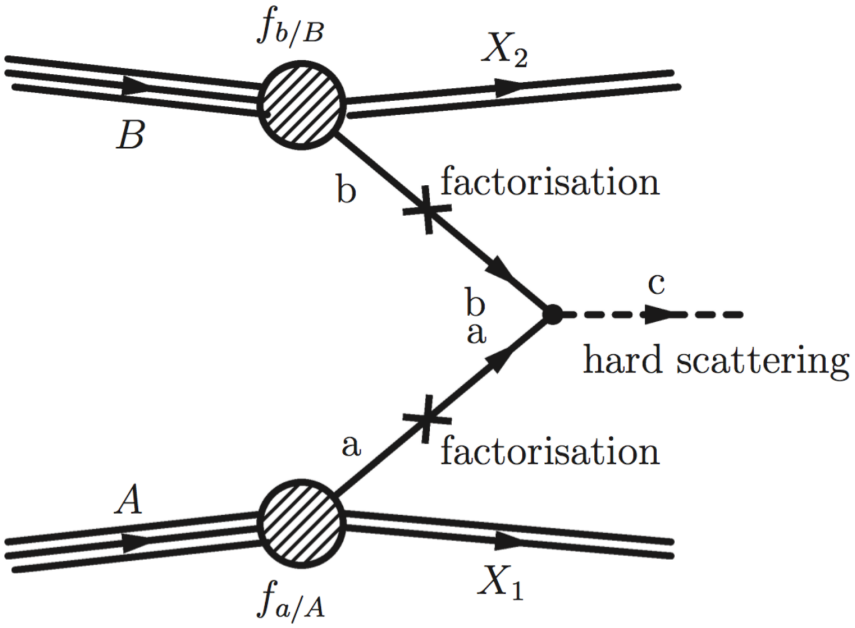


Figure 4.1: Hard scattering of a parton a from hadron A and parton b from hadron B to some final state c. The Figure is taken from [82].

4.2 PARTON SHOWERING AND HADRONIZATION

The ingoing and outgoing partons of the hard scattering process emit initial- and final-state radiation (ISR and FSR), therefore reducing their own energy. For example, gluons and quarks can emit a gluon, or a gluon can split into two quarks. The radiated partons mostly carry a small fraction of the initial parton's momentum and are therefore emitted at small angles. The latter partons can emit other partons, and this process, called a parton shower, will continue until the partons reach the non-perturbative regime, at an energy of about 1 GeV. At this point, the hadronization starts to happen: colored partons are clustered into hadrons. The initial parton coming from the hard scattering will therefore be represented by a collimated spray of energetic hadrons, called a jet.

The most commonly used model to describe the hadronization is the Lund String Model [83], implemented in PYTHIA [84] and proposed to be process independent. This model is based on the observation that the color potential of two quarks increases linearly with the distance between them [85]. When these two quarks, q and $\bar{q}$, move apart from their production vertex the potential will rapidly grow. As a result it will become energetically preferable to create a new quark-antiquark, $q\prime$ and $\bar{q}\prime$, pair between those partons, screening the color charge of the initial partons. The initial color singlet is now replaced by two $q\bar{q}\prime$ and $q\prime\bar{q}$ independent color singlets. This process repeats until the invariant masses of the color pairs are of the order of a hadron mass.

Apart from the hard interaction, other constituents of the colliding proton can also interact, giving additional hadrons in the final state. These interactions are called multiparton interactions. Beside this, the initial and final state gluon radiations not connected to the hard collisions and the multiparton interactions can happen. These are referred to as the underlying event.

4.3 PARTICLE PHYSICS GENERATORS

In order to compare the experimental observations to the theoretical predictions of the SM, the Monte Carlo (MC) event generators are used to simulate the physics processes.

MADGRAPH5_aMC@NLO [86] is a general purpose generator used for the computations of cross sections and generation of hard scattering process events by the calculation of ME for SM and BSM processes. These processes can be generated at Leading Order (LO) and next-to-leading-order(NLO) accuracy in the case of QCD corrections to SM processes; NLO refers to adding contributions with one extra loop to the process. For completing the event simulation after the ME calculation, an interface program, such as PYTHIA, is used for parton showering, hadronization and secondary interactions. The emission of an additional parton can now be obtained in two ways: either the ME includes the additional parton emission, or it could be emitted by the parton shower. This results in an overlap in the phase space and is called a "double-counting". To avoid this double-counting, a proper matching between the ME and parton shower is needed. In our studies, the MADGRAPH5_aMC@NLO samples are matched to PYTHIA using the MLM merging scheme [87], which is used for the processes generated at LO, and the FxFx merging scheme [88] for the processes generated at NLO.

POWHEG [89] is a dedicated NLO event generator that provides the modeling of the hard interaction at NLO QCD. It also needs then to be interfaced to PYTHIA for the parton showering and hadronization.

PYTHIA is an event generator, which can not only calculate the hard scattering process, but also has the tools needed for the calculation of the ISR and FSR parton showering,

hadronization, hadron decays, and the contribution from multiparton interactions and the underlying event. It can describe the complete final state of the event at the hadron level, but can provide only a limited set of ME processes.

The sketch of a hadron collision with all the effects discussed earlier and simulated in a MC generator is shown in Fig. 4.2.

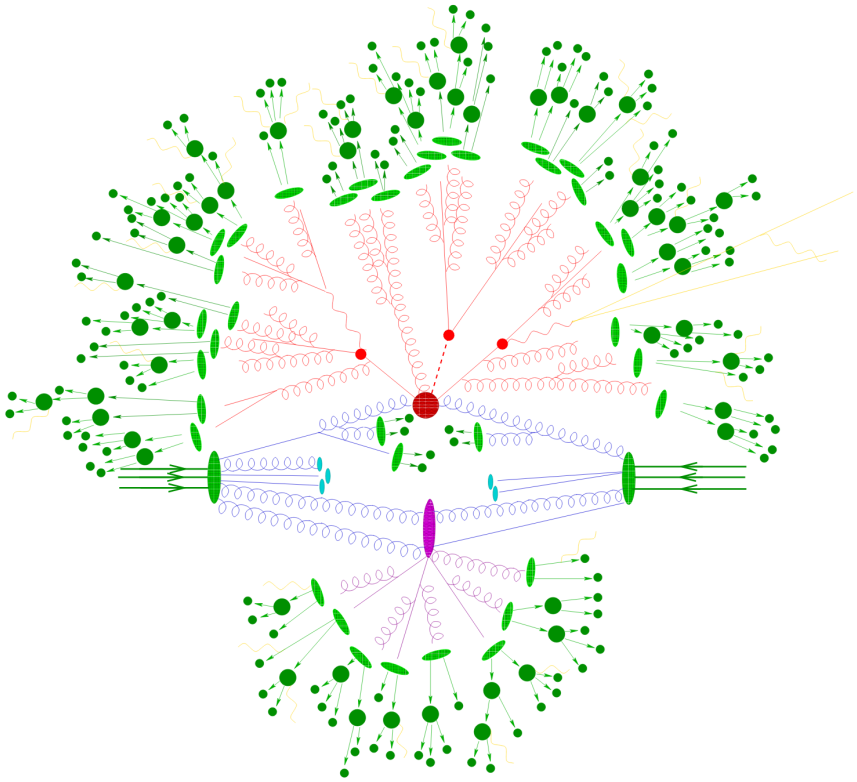


Figure 4.2: Sketch of a hadron-hadron collision as simulated by a MC event generator. The red blob in the center represents the hard collision, surrounded by a tree-like structure representing the parton showers. The purple blob indicates a secondary hard scattering event. Parton-to-hadron transitions are represented by light green blobs, dark green blobs indicate hadron decays, while yellow lines signal soft photon radiation. The Figure is taken from [90]

4.4 EVENT SIMULATION

The particles generated in MC events are transported into a full detector simulation, in order to have a description in terms of detector signals. The GEANT4 toolkit [91] is used to describe the full CMS detector geometry and is able to simulate the interaction of the particles with material of the sub-detectors. Each particle in the event is propagated through the different detector layers, including the dead material regions due to cables and support structures. In addition, secondary particles are generated originating from the interactions between the particles and the detector material. For each sub-detector, the response is accurately simulated providing an output signal in the same format as it appears in real data. CMS uses two types of MC event simulations - fast, the so-called "FastSim", and full, the so-called "FullSim", depending on analysis requirements. The "FastSim" method [92] reduces the CPU time consumption and is much faster compared to the "FullSim". This is extremely useful for physics analyses that require generation

of many samples in order to cover a wide phase-space region of physics parameters, i.e. searches for SUSY particles. While "FastSim" uses a parametric approach to simulate and reconstruct events, the "FullSim" approach is more time-consuming, but more accurate.

The CMS simulation uses minimum-bias[1] events created by an event generator (e.g., PYTHIA) to simulate the underlying interactions due to peripheral proton-proton collisions (PU) in each LHC bunch crossing. Due to electronics shaping time exceeding 25ns for some subdetectors, several bunch crossings before and after the bunch-crossing of interest must be included in the simulation, leading to hundreds of minimum bias events being used for each simulated hard-scattering event. At the next simulation step, information is converted into electronic signals including electronic noise. The information from first level trigger and HLT is also included in the simulated MC samples. The events in data and simulation are studied using the same reconstruction methods, allowing for a direct comparison of objects at the detector level.

1 By minimum-bias interaction any interaction that produce some detectable particles is meant

5

EVENT RECONSTRUCTION

This thesis is devoted to the searches for SUSY particles and cross section measurements for $t\bar{t}W$ and $t\bar{t}Z$ processes in multilepton final state. This requires excellent identification of at least two leptons, as well as light and heavy-flavor jets in case of strong SUSY production and $t\bar{t}V$ processes. The purpose of this chapter is to describe techniques and methods necessary to reconstruct these physics objects.

The event reconstruction in CMS detector provides an event description in terms of detected electrons, muons, photons, charged and neutral hadrons. These particles, produced in a hard scattering process, traverse the CMS sub-detectors and induce detector signals; a schematic view is shown in Fig. 5.1. Electrons and muons give different signatures, i.e. electrons produce hits in the tracker and are absorbed in the ECAL, while muons traverse all CMS sub-detectors and are registered in the Muon Chambers. Charged hadrons, such as charged pions, produce hits in the tracker, ECAL and HCAL, while neutral particles as photons and neutral hadrons do not produce any signals in the tracker, while being absorbed in ECAL and HCAL, respectively.

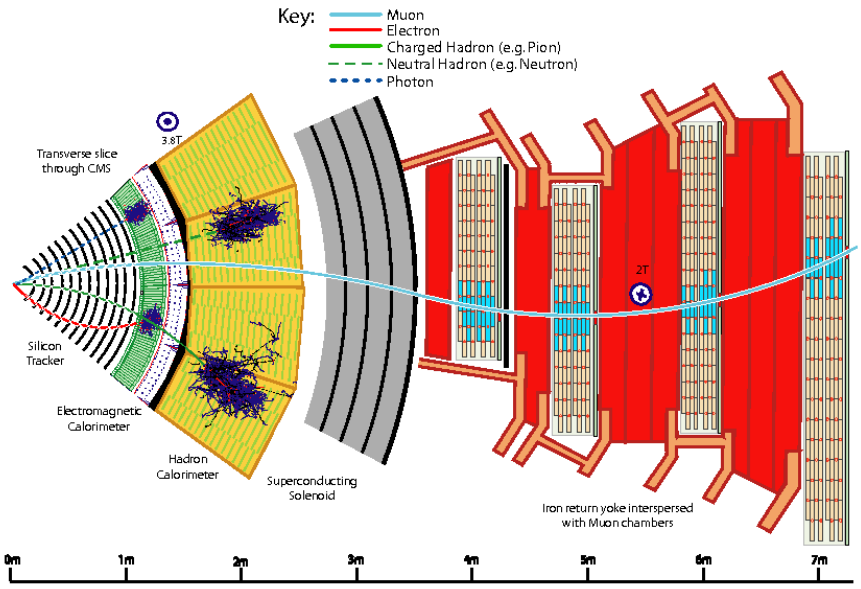


Figure 5.1: The slice of the CMS detector in x-y plane showing the trajectories of particles and their hits in the different sub-detectors. A track can be reconstructed for charged particles, such as electrons, muons and charged hadrons. Electrons and photons deposit their energy in the ECAL, whereas hadrons are stopped in the HCAL. The muons transverse all sub-detectors and are stopped only in the Muon Chambers. The Figure is taken from [93].

A technique called Particle Flow (PF) [93] is used to reconstruct and identify all stable particles in the event[1] by combining the signals from different sub-detectors. Neutrinos do not leave a trace in the detector, but still can be revealed as the missing energy in the transverse plane (p_{T}^{miss}). More complicated identification criteria are usually required in order to reduce the misidentification rate. The event reconstruction and additional identification techniques are applied in the same way to data and simulation. To account for a remaining mismodeling, event reweighting techniques are applied to MC simulation. We will briefly discuss it in this chapter.

5.1 COMPONENTS OF THE PARTICLE FLOW ALGORITHM

The separate components of the PF algorithm are explained below and includes tracks and vertex reconstruction, clustering in the calorimeters, the linking algorithm between the tracks and clusters and as well the discussion on how PF is used to suppress PU effects.

5.1.1 *Tracks and vertex reconstruction*

The tracker provides a very good momentum resolution for charged hadrons, better than the calorimeters, up to several hundred GeV. It also gives a precise measurement of the direction of charged particles. For these reasons, the tracker is the cornerstone of the PF algorithm.

Track reconstruction in CMS [68] is based on the very stringent criteria for the track seeds and reconstruction quality. Each iteration starts with a seed using only a few hits and providing information on initial track parameters, i.e. the position and direction vectors and an initial estimate of the transverse momentum. The hits are chosen in the innermost layers of tracker as the high granularity of the pixel detectors ensures lower occupancy and provides a better estimate for the initial parameters. The initial track parameters are extrapolated along the expected flight path of a charged particle, searching for additional hits along the trajectory. The Combinatorial Kalman Filter [94] is used for this search: the track parameters at each detector layer are used to find a compatible measurement in the next detector layer. If it is found, the track parameters are updated taking into account this information before extrapolating to the subsequent layer. The hits assigned to the tracks are then removed, and the tracking cycle is repeated several times for the remaining hits with progressively looser track seeding criteria. The looser criteria increase the tracking efficiency, and the misidentification rate is kept low because of the reduced combinatorics resulting from the removal of hits in the previous iterations. During additional iterations, the constraints on the vertex origin are also relaxed, allowing to reconstruct the tracks associated with secondary vertexes.

At least one vertex is required to be reconstructed in each event. This vertex should be associated with at least four charged-particle tracks. It should also lie within 24cm of the origin of the CMS coordinate system along the beam direction, $\mathrm{d_z}$, and within 2cm in the plane transverse to the beam, $\mathrm{d_{xy}}$[2]. In case multiple vertexes are reconstructed in an event, the vertex with the highest value for a sum of p_T^2 of the associated tracks and squared missing transverse momentum, estimated with these tracks, is chosen to be the leading primary vertex in the event. All other vertexes are referred to as PU vertexes.

1 these particles are stable with respect to the time needed to traverse the detector
2 both $\mathrm{d_{xy}}$ and $\mathrm{d_z}$ will be referred to as impact parameter variables

5.1.2 *Calorimeter clustering*

The clustering algorithm starts with seeds which are found as a local maximum above a given energy threshold. From the seed, the clusters are grown by aggregating cells which exceed a given energy threshold and have at least one side in common with a cell already in the cluster. The thresholds are taken at two standard deviations above the electronics noise level in the calorimeter. For each seed within the topological clusters, exactly one PF cluster is constructed. The energy and position of each PF cluster are determined using an iterative procedure, starting from the seed positions as an initial estimate for the PF cluster positions. The energy of each cell within the topological cluster is shared between the different PF clusters based on the distance between the cell and the cluster. The clustering is performed separately for the barrel and endcap sub-detectors in the ECAL and HCAL, and for the pre-shower, while the clustering is not performed for the forward calorimeter where each cell gives rise to exactly one cluster.

5.1.3 *Linking algorithm*

The tracks and calorimeter clusters need to be connected to each other by a linking algorithm to reconstruct each single particle, while getting rid of any possible double counting from different sub-detectors. Tracks are extrapolated from their last measured hit in the tracker to the calorimeters, where they are linked to any given cluster if the extrapolated position is within the cluster boundaries. Links between calorimeters are established when the cluster position in a higher granularity calorimeter is within the cluster boundaries of a less granular calorimeter. The distance in the η-ϕ plane between two linked elements gives a quantified description for the quality of the matching. Links between tracks reconstructed individually in the silicon tracker and in the muon system are established when the global fit between two tracks returns an acceptable χ^2 value. When several tracks from the silicon tracker can be linked to a muon track, only the track with the smallest χ^2 is picked.

5.1.4 *Particle flow event reconstruction*

Elements which are directly or indirectly linked are grouped together in blocks. Due to the fine granularity in the CMS sub-detectors, each block typically only contains few elements. For complex events, the number of blocks will increase, while the number of elements in each block remains the same, making the performance of the algorithm independent from the event complexity. The blocks are used as simple inputs to build the list of individual particles. First, each global muon is added to the list of PF muons if its momentum is compatible with the momentum determined using tracker information only. The track is then removed from the block.

A Gaussian-Sum Filter (GSF) [95] is used to refit the candidate electron track, taking into account the possible energy loss by bremsstrahlung, and follow its trajectory to the ECAL. Seeds for the time-consuming fitting procedure are chosen only from the subset of tracks that pass certain identification criteria. An electron is fully identified if its track is matched with an ECAL cluster, and if it passes a set of tracking and calorimeter requirements. The electron is then added to the PF electron collection, and the associated electron track and ECAL clusters, including those from bremsstrahlung, are removed from the block before further processing.

The remaining tracks are subject to tighter quality criteria: tracks for which the relative uncertainty on the p_T is smaller than the expected relative calorimeter energy resolution for charged hadrons are rejected. Each of the remaining tracks give rise to a PF charged hadron. If the energy in the calorimeters is compatible with the track momentum within uncertainties, the charged hadron momentum is redefined by a fit of the measurements in the tracker and the calorimeter. It is also possible that the calibrated energy of the closest ECAL and HCAL clusters, linked to the track, is significantly larger compared to what is expected from the momenta in the tracker. If this excess is larger than the expected calorimeter energy resolution, a PF photon or PF neutral hadron is defined. In addition calorimeter clusters, which are not linked to any track, give rise to PF photons or PF neutral hadrons. At the end of the PF event reconstruction procedure, we have a list of identified particle objects.

5.1.5 *Pileup mitigation techniques*

The presence of PU causes extra energy deposits in calorimeters and tracks which are overlaid with a hard interaction event. This leads to degradation in resolution and less clean signatures. PU vertexes are usually separated in space from the vertex of interest. The CMS precise tracker system allows these vertexes to be reconstructed, and we can use the PF framework to mitigate the effect of PU, using a technique, the so-called charged hadron subtraction [96].

Contamination from PU events is reduced by discarding charged hadron PF candidates that are associated to PU vertexes, before any jet clustering and any further processing. Charged hadron candidates with a track associated to a PU vertex are removed. All other tracks are retained.

5.2 PHYSICS OBJECT IDENTIFICATION

The event selection is an integral part of any physics analysis. It determines which events are used, and thus which processes contribute to the specific dataset. An event is usually selected based on several physics objects, e.g. electrons, muons and jets[3], which are decay products of the initially produced unstable particles. The particle flow technique described in the previous section is very compatible with this approach, given that it reconstructs a fully consistent set of identified particles from the detector hits.

However, a more careful selection of the PF objects is needed in order to ensure that their behavior is well understood, and to ensure that the selected events are not dominated by misidentified particles, or due to detector effects. The physics object groups (POG's) within the CMS Collaboration are in charge of providing general recommendations on how to define each object. These recommendations are based on extensive studies, and are applicable for most analyses, thus reducing the workload for the analysis teams. However, it is also possible to develop a technique which is suitable for particular analysis, but then this technique should be approved internally within CMS POG groups. In the following, we describe quality criteria for selecting electrons, muons, jets and missing transverse energy for physics analysis.

3 Further in the thesis charged lepton is denoted as ℓ, jet as j and b-tagged jet as b

5.2.1 *Muons*

In addition to the tracks reconstructed in the tracker, CMS uses standalone muon identification based on the muon chambers. Hits in DTs and CSCs are geometrically matched to form track segments. The track segments in the innermost muon chambers are used as seeds to built a standalone-muon track towards the outer layers using the Kalman Filter technique. Muons are reconstructed in CMS using two approaches [97]: global-muon and tracker-muon.

Global-muon reconstruction starts from a standalone-muon track matched to a track in the inner tracker; tracks from both systems are combined in a global-muon track using a Kalman Filter to update parameters. For large transverse momenta, the global-muon fit improves the resolution compared to the tracker-only fit.

Tracker-muon reconstruction where inner tracks are extrapolated to the muon system, taking into account its expected trajectory through the magnetic field including average expected energy losses and multiple Coulomb scattering in the detector material. If the track can be matched to a muon segment, the track qualifies as a tracker-muon. In general, the tracker muon reconstruction is more efficient for muons with low momenta.

The two working points, so called "Loose ID" and "Medium ID", are used for the muon identification. Only muons within the muon system acceptance $|\eta| < 2.4$ and a minimum p_{T} of 10 GeV are considered. "Loose ID" aims to identify muons from W or Z boson prompt decay and muons from light and heavy flavor decay, while keeping a low rate of the misidentification of charged hadrons as muons. "Medium ID" is optimized for prompt muons and for muons from heavy flavor decay. For detailed explanation of identification criteria and optimization procedure for both "Loose ID " and "Medium ID" working points we refer the reader to the Ref. [97].

One additional identification variable used in the analyses is the muon segment compatibility. This variable is calculated by propagating the inner track to the muon system, and evaluating both the number of matched segments in all stations and the closeness of the matching in position and direction [97].

5.2.2 *Electrons*

Electron candidates are reconstructed by associating trajectories in the inner tracking system with energy deposits in the ECAL [95]. The geometrical acceptance is restricted by the tracking system to $|\eta| < 2.5$. Two complementary approaches for seeding electron candidates are employed. The first one uses clustered energy deposits in the ECAL as a starting point to predict the position of hits in the tracking detector, the second one matches charged particle tracks with clusters in the ECAL. The clustering algorithms take into account that electrons can radiate large fractions of their energy as bremsstrahlung when passing through the detector material. This causes their energy to be spread over an array of ECAL crystals. For the track reconstruction, this energy loss leads to the increase of the curvature of the track trajectory and requires to employ a dedicated Gaussian sum filter [95].

The electron seeds are not always the result of isolated electrons from a primary vertex, but might be due to background processes, mainly originating from photon conversions, or from jets, misidentified as electrons. Electron identification algorithm is used to discriminate between prompt and nonprompt[4] electrons, the used variables can be categorized into three classes:

4 Hereafter nonprompt leptons will be referred to all leptons not originated from W or Z boson

- observables based on the shape of the electromagnetic cluster, e.g., the width of the cluster along the η direction
- observables based on tracking information, e.g., the momentum loss due to bremsstrahlung
- the quality of the matching between the supercluster and the track, e.g., the ratio of the supercluster energy over the track momentum

The electron identification is based on a Boosted Decision Tree (BDT) that combines 20 variables from these categories. The full list of variables and a description can be found in Table 5.1. The BDT returns a real number (score) between -1 and 1 for each electron candidate, where large positive values correspond to signal-like hypothesis. The training is done in three distinct regions of η, namely in the inner barrel ($|\eta| < 0.8$), outer barrel ($0.8 < |\eta| < 1.479$), and endcaps ($1.479 < |\eta| < 2.5$).

Observable type	Observable	Definition
cluster shape	$\sigma_{i\eta i\eta}$, $\sigma_{i\varphi i\varphi}$	Standard deviation of the energy distribution in the cluster along the η, φ direction of the energy-crystal number spectrum along η and φ
	$\Delta\eta_{SC}$	Super cluster width along η
	$\Delta\varphi_{SC}$	Super cluster width along φ
	H/E	Ratio of the hadronic energy behind the electron supercluster to the supercluster energy
	$(E_{5\times5} - E_{5\times1})/E_{5\times5}$	Circularity: the energy sums $E_{i\times j}$ of the i crystals in φ and j crystals in η centred on the seed crystal
	$R_9 = E_{3\times3}/E_{SC}$	Ratio of the energy in a 3×3 (9 crystal) cluster around the seed over the SC energy
	E_{PS}/E_{raw}	For endcap training bins only: energy fraction in pre-shower over the raw SC energy
tracking	$f_{brem} = 1 - p_{out}/p_{in}$	Fractional momentum loss as measured by the GSF fit. The momenta p_{in} and p_{out} are extrapolations of the GSF track to the vertex and ECAL respectively.
	N_{KF}	Number of hits of the Kalman Filter track of the iterative combinatorial track finder, if any
	χ^2_{KF}	Reduced χ^2 of the KF track
track-cluster matching	E_{SC}/p_{in}	Ratio of the SC energy and the track momentum at the innermost hit
	E_{ele}/p_{out}	Ratio of the energy of the cluster closest to the electron track and the track momentum at the outermost hit
	$1/E_{tot} - 1/p_{in}$	Energy-momentum agreement
	$\Delta\eta_{in} = \lvert\eta_{SC} - \eta_{in}\rvert$	Distance between the energy-weighted center of the SC and the expected shower position as extrapolated from the GSF trajectory state at the vertex
	$\Delta\varphi_{in}$	Same as $\Delta\eta_{in}$ but along φ
	$\Delta\eta_{seed} = \lvert\eta_{seed} - \eta_{out}\rvert$	Distance between the pseudorapidity of the seed cluster and the expected shower position as extrapolated from the GSF trajectory state of the outermost hit

Table 5.1: Overview of input variables used for electron multivariate identification. The table is taken from [98].

Additional requirements are applied for all selected electrons to ensure that the offline selection is more stringent than the selection at trigger level. These are variables describing the shower shape, the ratio of energy deposits in the ECAL and HCAL, and the relation of energy E and momentum of an electron ($|1/E - 1/p|$). In the following, we refer to them as the "trigger emulation" requirement.

To improve the accuracy of the electron charge reconstruction, the position of the calorimeter deposit, identified by the linear projection of the deposits in the pixel detector to the inner calorimeter surface, has to match with a position determined with the full track fit. We refer to this requirement as the "charge consistency" requirement.

Another requirement used in the analyses is linked to the elimination of electrons originating from photon conversions. The contribution from these electrons is suppressed by rejecting candidates that do not produce energy deposits in the innermost tracking layers, or which are associated with a displaced vertex compatible with a photon conversion. This requirement will further be referred to as "conversion rejection".

5.2.3 *Isolation and additional identification criteria*

Another set of variables allows to discriminate between prompt and nonprompt leptons. The isolation is a powerful tool to reduce backgrounds coming from hadronic jets; prompt leptons are usually not surrounded by other particles coming from the hard interaction. Therefore, they tend to be isolated from other particles in the event. In practice one sums the p_{T} of all reconstructed PF candidates around the lepton. This is done separately for charged hadrons, neutral hadrons and photons. The lack of a reliable vertex information for photons and neutral hadrons makes the latter two contributions susceptible to the effect of PU in the event and a correction needs to be applied to reflect the contribution from neutral particles. The relative isolation is defined as:

$$I_{rel} = \frac{\sum_R p_{\mathrm{T}}(h^{\pm}) - \max(0, \sum_R p_{\mathrm{T}}(h^0) + p_{\mathrm{T}}(\gamma) - p_{\mathrm{T}}^{\mathrm{PU}})}{p_{\mathrm{T}}(\ell)} \tag{5.1}$$

where the sum runs over all PF candidates within the cone size of R. Usually for standard isolation two values are considered, 0.3 and 0.4. Charged-hadron candidates associated with PU vertexes are not taken into account in the calculation of the isolation. However, they can be used to estimate the remaining contribution to the isolation coming from neutral hadrons associated with PU. This type of correction is referred to as $\Delta\beta$ correction and is calculated with a formula:

$$p_{\mathrm{T}}^{\mathrm{PU}} = 0.5 \sum_R p_{\mathrm{T}}^{charged,\mathrm{PU}} \tag{5.2}$$

The factor 0.5 has been measured in jets [99] and empirically found that the total energy coming from neutral hadrons and photons is on average one half of the energy of the charged hadrons originating from PU.

Another approach for PU correction exploits an estimate of the median energy from PU as calculated with the FastJet algorithm [100]. The PU correction in this case is based on the effective area (EA) technique and estimated as:

$$p_{\mathrm{T}}^{\mathrm{PU}} = \rho A_{eff}, \tag{5.3}$$

where ρ is the mean energy density in the event and the effective area A_{eff} is determined in bins of $|\eta|$ and defined as the ratio between the slope of the average isolation and that of ρ as a function of the number of reconstructed vertexes [98]. The values for effective area are determined separately for each flavor and each dataset and can be found in Ref. [101] and [102].

The distinction between prompt and nonprompt becomes less evident in systems with a high Lorentz boost, where decay products tend to overlap and jets may contribute to the energy deposition around prompt leptons. This problem is mitigated with additional variables. Among them, mini-isolation (I_{mini}) is defined as the relative isolation, but the cone around the lepton has a p_{T}^{ℓ}-dependent radius [103] according to the following formula:

$$R = \frac{10\,\mathrm{GeV}}{\min(\max(p_{\mathrm{T}}^{\ell}, 50\,\mathrm{GeV}), 200\,\mathrm{GeV})}. \tag{5.4}$$

A downside of using only mini-isolation is that certain nonprompt leptons can pass the isolation requirement, especially if their p_{T}^{ℓ} is large enough to decrease the cone size.

Also this can be the case for nonprompt leptons that originate from the decay of low-p_{T} b quarks. For such decays, the angle between the trajectory of the lepton and the jet that arises from the hadronization of the light quark can exceed the opening angle of the isolation cone, which causes them to pass the isolation requirement. Better rejection of nonprompt leptons can be achieved by using another two variables: $p_{\mathrm{T}}^{\mathrm{ratio}}$ and $p_{\mathrm{T}}^{\mathrm{rel}}$.

The $p_{\mathrm{T}}^{\mathrm{ratio}}$ variable is defined as the ratio of the lepton p_{T} and of the jet, matched to the lepton. This jet is matched geometrically to the lepton within $\Delta R = 0.4$. If no jet is matching the lepton, then the $p_{\mathrm{T}}^{\mathrm{ratio}}$ is equal to 1. The use of $p_{\mathrm{T}}^{\mathrm{ratio}}$ is a simple way to identify leptons in quite boosted topologies, without any jet reclustering. The $p_{\mathrm{T}}^{\mathrm{rel}}$ variable is defined as the magnitude of the component of the lepton momentum perpendicular to the axis of the closest jet:

$$p_{\mathrm{T}}^{\mathrm{rel}} = \frac{||(\vec{p}(\mathrm{jet}) - \vec{p}(\ell)) \times \vec{p}(\ell)||}{||\vec{p}(\mathrm{jet}) - \vec{p}(\ell)||}. \tag{5.5}$$

If no jet is matching the lepton, the $p_{\mathrm{T}}^{\mathrm{rel}}$ variable is set to 0. This variable allows recovering leptons from accidental overlap with jets in Lorentz-boosted topologies. For the calculation of $p_{\mathrm{T}}^{\mathrm{rel}}$ and the previously mentioned $p_{\mathrm{T}}^{\mathrm{ratio}}$, jets with $p_{\mathrm{T}} > 10$ GeV and without any additional identification criteria are considered.

A powerful discrimination between nonprompt and prompt leptons for SUSY searches in multilepton final state can be achieved by requiring that a lepton candidate is locally isolated (small I_{mini}) and either carries the major part of the momentum of the associated jet (large $p_{\mathrm{T}}^{\mathrm{ratio}}$) or is considered to overlap with the jet only accidentally (large $p_{\mathrm{T}}^{\mathrm{rel}}$) and will be referred to as the multi-isolation requirement. It can be formulated as:

$$(I_{\mathrm{mini}} < I_{\mathrm{mini}}^{\mathrm{cut}})\,\mathrm{AND}\,(p_{\mathrm{T}}^{\mathrm{ratio}} > p_{\mathrm{T}}^{\mathrm{ratio,\,cut}}\,\mathrm{OR}\,p_{\mathrm{T}}^{\mathrm{rel}} > p_{\mathrm{T}}^{\mathrm{rel,\,cut}}), \tag{5.6}$$

where $I_{\mathrm{mini}}^{\mathrm{cut}}$, $p_{\mathrm{T}}^{\mathrm{ratio,\,cut}}$ and $p_{\mathrm{T}}^{\mathrm{rel,\,cut}}$ are criteria tuned to define working points with different prompt lepton efficiencies and nonprompt lepton misidentification rates. These criteria are flavor dependent and optimized to achieve comparable selection efficiencies for prompt electrons and muons. Owing to the larger probability of electron misidentification compared to muons, this results in a more stringent requirements for electrons, which is shown in Table 5.2.

	Muons	Electrons
$I_{\mathrm{mini}}^{\mathrm{cut}}$	0.16	0.12
$p_{\mathrm{T}}^{\mathrm{ratio,\,cut}}$	0.69	0.76
$p_{\mathrm{T}}^{\mathrm{rel,\,cut}}$	6.0	7.2

Table 5.2: Flavor dependent requirements on the isolation variables I_{mini}, $p_{\mathrm{T}}^{\mathrm{ratio}}$ and $p_{\mathrm{T}}^{\mathrm{rel}}$ employed in the SUSY analysis lepton selection.

MULTIVARIATE DISCRIMINANT FOR PROMPT LEPTON SELECTION A more robust algorithm to separate between prompt and nonprompt leptons is the ultimate, multivariate discriminant referred to as Lepton MVA. A number of discriminating variables is used; among them: lepton p_{T} and η, the impact parameters of the leptons,

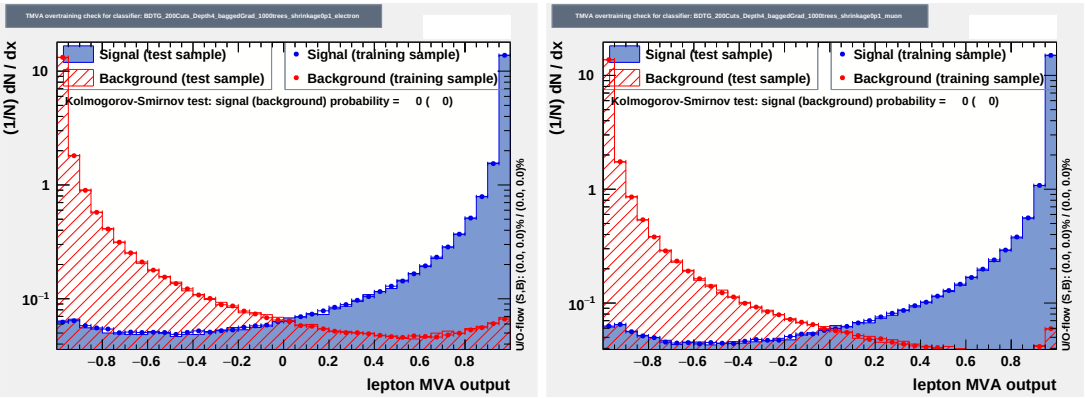


Figure 5.2: Lepton MVA output shape for prompt, labeled signal, and nonprompt, labeled background, electrons (left) and muons (right). The figures are made by Willem Verbeke.

and their significance, $\mathrm{SIP_{3D}}$[5] ensures that leptons originate from the primary interaction vertex, relative isolation in cone R = 0.3, and mini-isolation I_{mini}. The discriminator also exploits information about the closest jet to the lepton, therefore $p_{\mathrm{T}}^{\mathrm{ratio}}$, $p_{\mathrm{T}}^{\mathrm{rel}}$, the number of charged particle tracks in the jet and the jet's value of Deep CSV b tagging discriminator, which will be explained later, also serve as inputs to the Lepton MVA discriminator. In addition to all aforementioned variables, the muon segment compatibility is used in muon selection, and the electron MVA in the selection of electrons.

All variables are combined into a BDT discriminant, which is trained and evaluated with the TMVA package [104]. The training is done using simulated prompt leptons originating from the tZq and $\mathrm{t\bar{t}Z}$ processes, while nonprompt leptons were taken from $\mathrm{t\bar{t}}$ simulation samples. Individual lepton MVA discriminants are trained separately for MC simulation in 2016 and 2017 campaigns, and separately for muons and electrons. The leptons are preselected using quite loose requirement on impact parameters and isolation, and muons were additionally required to pass the "Medium ID" requirement. All in all, close to 15 million prompt, and 2.5 million nonprompt leptons were available for each of the four lepton MVA training iterations and 80% of these were used to train the lepton MVA, while the remaining 20% were used as a test sample. For details of the training we refer the reader to the Ref. [105].

The output of the BDT training for the 2016 data sample is shown in Fig. 5.2 for prompt and nonprompt electrons and muons. The output distributions for training and test datasets are nearly identical, confirming that no significant effects of overfitting are present in the lepton MVA. The distributions for 2017 conditions look similar.

Receiver operating characteristic (ROC) curves, showing the signal efficiency as function of the background efficiency, for the lepton MVA, compared to several other identification criteria are shown in Fig. 5.3. The superior performance of the new lepton MVA for selecting prompt leptons, compared to selection criterion provided by CMS POG, and other multivariate lepton discriminants such as the one used in the recent $\mathrm{t\bar{t}H}$ discovery [36], is clearly visible. To make sure none of the discriminating variables fed into the lepton MVA discriminant show major mismodeling in data, all variables, including the lepton MVA value, were checked in several dilepton control regions and are shown in Appendix A.

5 the significant is defined as the ratio $\frac{\mathrm{d_{3D}}}{\sigma(\mathrm{d_{3D}})}$, where $\mathrm{d_{3D}}$ is the three-dimensional impact parameter with respect to the primary vertex and σ_{3D} is its uncertainty from the track fit

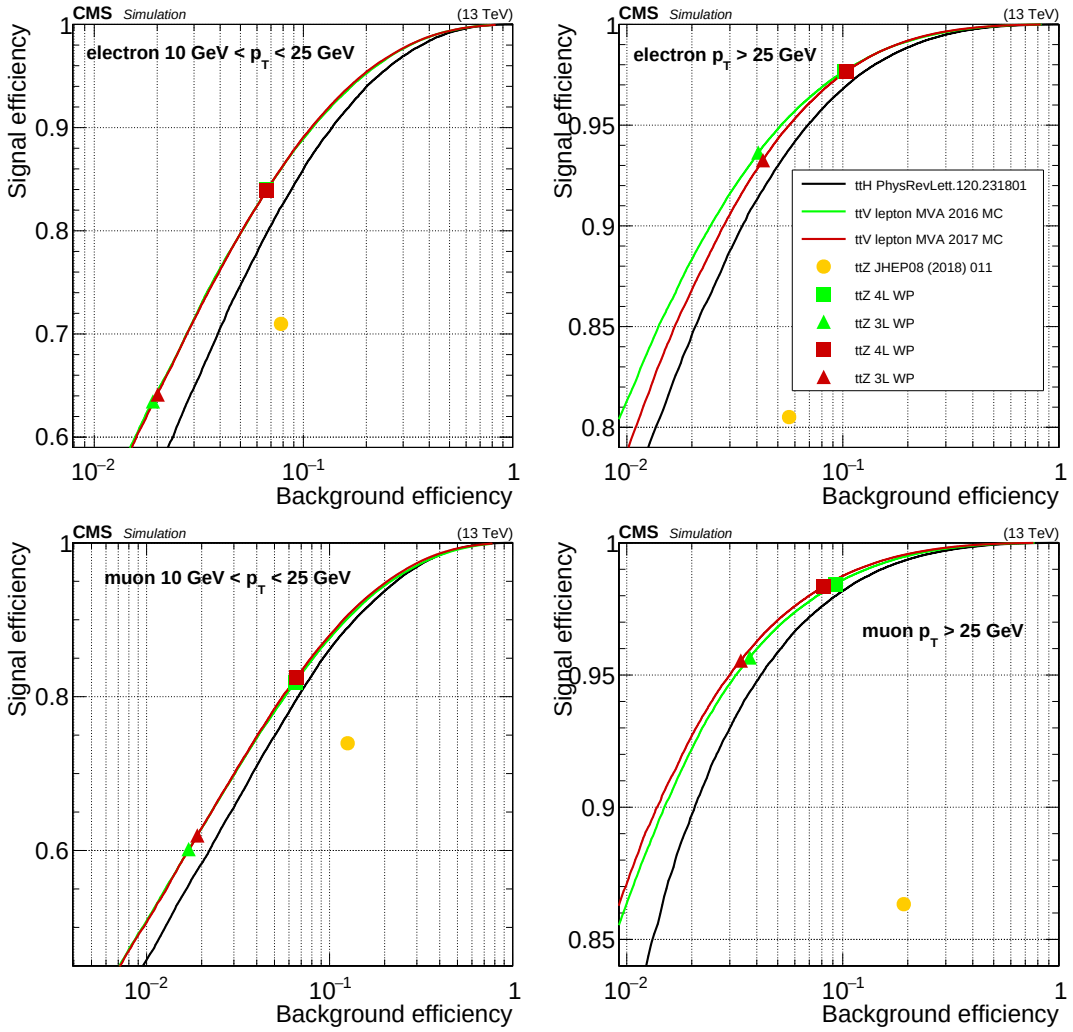


Figure 5.3: ROC curves for different lepton identification variables, including previous lepton MVA training used in $t\bar{t}H$ discovery [36] and the cut-based POG identification working point used in previous $t\bar{t}Z$ inclusive cross section measurement [45]. The top-left and top-right plots show the ROC curves for respectively low and high p_T electrons, while the bottom plots are concerned with low and high muon p_T.

5.2.4 *Jets*

Individual particles identified by the PF algorithm are clustered into jets if they originate from hadronization of the same parton in the hard interaction or in subsequent decays. The clustering is performed with the "anti-k_t" algorithm [106] using a distance parameter of R = 0.4. The "anti-k_t" is a sequential algorithm used as default in CMS experiment. It is infrared and collinear safe compared to cone algorithms used before LHC [107]. The "anti-k_t"" algorithm, as other sequential algorithms, is based on exploiting a quantity, which describes how likely two partons are stem from the same QCD splitting, and proceed sequentially to construct the jet by reconstructing the partons which are closer to each other in terms of this quantity. This is done in the following way. In first step two distances are defined: the distance between two objects i and j, d_{ij}, and the distance between the object i and the beam , d_{iB}, which are calculated according to

$$d_{ij} = \min(\frac{1}{p_{Ti}^2}, \frac{1}{p_{Tj}^2})\frac{\Delta R_{ij}^2}{R^2},$$
$$d_{iB} = \frac{1}{p_{Ti}^2}.$$

Here ΔR_{ij} is the cone width between two objects and R represents the cone width used for clustering. If d_{ij} is the smallest among two distances then i and j entities are combined into a new entity k, for which the four-momentum vector is defined as a sum of four vectors for i and j. In the other case, distance d_{iB} is assigned to a jet with index i and and i's object is removed from the list of entities and considered as a final jet. The algorithm is then repeated until no entities are left in the event.

The functionality of the algorithm can be understood by considering several particle combinations. The distance is inversely proportional with the transverse momentum such that particles with high momentum (hard hadrons) will be given priority in the clustering and hard hadrons will tend to cluster with soft (low momenta) hadrons. This is illustrated in Fig. 5.4. The algorithm yields circular jets around high-p_T particles, whose shapes are resilient to nearby low-p_T particles originating for example from PU. The shapes of adjacent low-p_T jets, however, can have non-circular shapes if they overlap with the hard jet. To reject noise and to avoid mismeasurements, PF jets considered for this analysis have to meet additional requirements.

JET IDENTIFICATION Jet identification criteria were applied in order to reject fake jets originating from calorimeter and readout electronics noise, while retaining the vast majority of the jets ($> 99\%$). Jets passing the "Loose" ("Tight") working point for jet identification used in analyses exploiting $2015 - 2016$ (2017) datasets, the details of selection are shown in Table 5.3.

Only jets above a transverse momentum $p_T > 30$ GeV and within the tracker acceptance $|\eta| < 2.4$ are considered. The jet reconstruction algorithm uses as input the entire list of PF objects including the leptons. Therefore the leptons originating from W and Z bosons are also potentially clustered as jets. These jets need to be removed from the jet list, which is achieved by vetoing jets that are within the $\Delta R < 0.4$ cone around the selected leptons. The number of selected jets in the event in particular analysis is denoted as N_j. From those selected jets, the key variable H_T is defined by:

$$H_T = \sum_{jets} p_T.$$

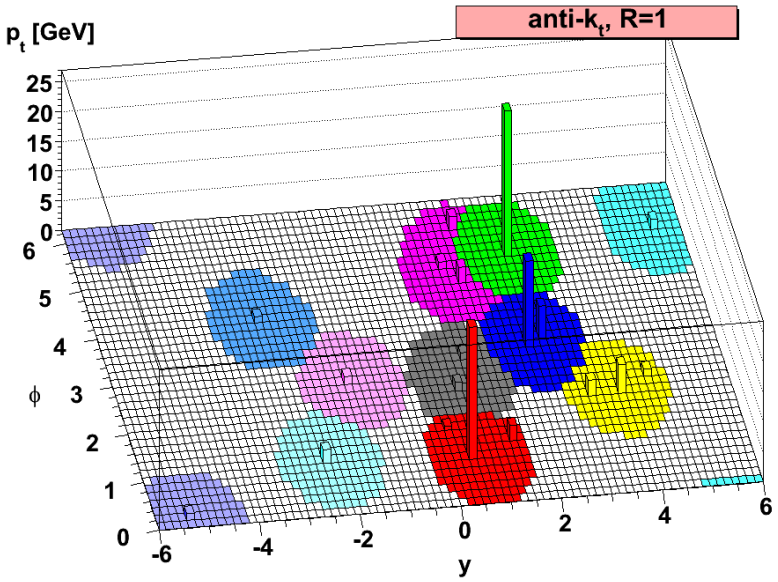


Figure 5.4: Illustration of jets obtained with the anti-k_t algorithm, showing the active areas in η-ϕ space which tend to be circular. The figure is taken from [106].

	Loose	Tight
Neutral hadronic energy fraction	< 0.99	< 0.9
Neutral electromagnetic energy fraction	< 0.99	< 0.9
Charged electromagnetic fraction	< 0.99	-
Number of constituents	> 1	> 1
Charged Multiplicity	> 0	> 0
Charged Hadron Fraction	> 0	> 0

Table 5.3: Summary of the selection applied for jet identification

JET ENERGY CORRECTIONS AND RESOLUTION In order to account for the contribution of PU to the measured jet energy and to compensate for nonlinear detector responses, jet energy corrections (JEC) [108, 109] are applied to relate the measured energy to the true energy deposit at the particle level. The overall correction factor is composed of the offset correction, the MC calibration factor, and the residual calibrations. Each correction factor is applied consequently.

At first, the offset correction is applied to subtract the energy not associated with the main primary vertex. The collisions taking place within a single beam crossing are called in-time PU. The contribution from the previous and subsequent beam crossings, which can also contribute to calorimetric energy in the same time window as the primary hard interaction, due to the finite signal decay time in calorimeters, are called out-of-time PU. The energy excess in the jet includes contributions from both in-time and out-of-time PU interactions, and from electronic noise in the calorimeters. The in-time PU is removed by identifying and subtracting charged hadrons as explained earlier and the remaining energy due to out-of-time PU and neutral particles is estimated with a effective area

approach, which is explained earlier for PU corrections in isolation estimation. This correction is applied both to MC and data.

After removing the offset contribution, the energy of the jet needs further corrections for the non-uniform and non-linear detector response. These effects are mostly mitigated by correction factors determined from studies on simulated jets in MC, in which the energy of the reconstructed jets were matched with the energy of their corresponding generator-level jets. In this way, scale factors are derived in bins of p_T and η.

On top of the simulation-based correction, small residual corrections in p_T and η are applied only on data. The response of all jets in the event are corrected relatively to the jets in the CMS barrel region. The barrel region is chosen as the reference region because the detector material there is more uniform. The corrections are determined using a high statistics sample with two jets, in which one jet is found in the barrel region and is used as the reference object, while the other jet is free to scan the whole detector. Because of momentum conservation, we can assume both jets to be balanced in the transverse plane. Deviations from this balance are used to derive corrections in bins of η.

The jet energy resolution also needs to be calibrated. Measurements show that the jet energy resolution (JER) in data is lower than the one in the simulation. Therefore, the jets in MC need to be smeared to match the jets from data.

HEAVY FLAVOR TAGGING The suppression of b quark decays by the Cabibbo-Kobayashi-Maskawa (CKM) matrix [110, 111] leads to comparatively long lifetimes of $O(10^{-12}$ s) for particles containing a b quark. This entails that the distance between the primary vertex and the point where such particles decay is usually long enough to be resolved with the CMS pixel detector. The resulting so-called b jets can therefore be discriminated against jets originating from light flavor quarks or gluons by detecting displaced secondary vertexes. This allows to enhance the sensitivity of the search for signal models that are expected to produce this signature by categorizing events according to the number of identified b jets.

The identification of b jets is performed with the combined secondary vertex algorithm (CSVv2) in the SUSY searches and the $t\bar{t}W$ cross section measurement [112]. By the time of the $t\bar{t}Z$ cross section measurement, the b tagging identification was improved using advances in deep machine learning [113] and a new version of the CSVv2 tagger was developed. The Deep CSV algorithm uses a deep neural network with more hidden layers, more nodes per layer, and a simultaneous training for all vertex categories and for all jet flavors. The efficiencies of the CSVv2 and Deep CSV algorithms can be compared in Fig. 5.5. The working point of the CSVv2 algorithm chosen for SUSY searches and the cross section measurement for the $t\bar{t}W$ analysis has an efficiency of 65% to correctly tag a jet as b jet, while keeping the mistagging rate less than 1%. The Deep CSV algorithm achieves a p_T averaged efficiency of 70% for b quark jets to be correctly identified as b jets while retaining a mistagging rate 12% for charm quark jets and 1% for jets originating from u, d, or s quarks or from gluons. The number of selected b jets in the event of a particular analysis is denoted as N_b.

5.2.5 *Missing Transverse Momentum*

The total transverse momentum of the particles that do not interact with the detector material is obtained by balancing the momentum conservation in the transverse plane. It

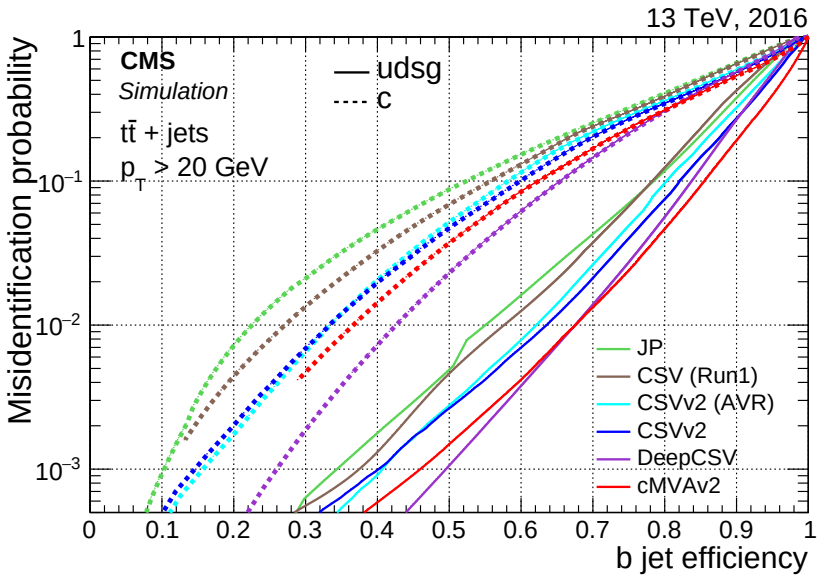


Figure 5.5: Misidentification probability for c and light-flavor jets versus b jet identification efficiency for various b tagging algorithms applied to jets in $t\bar{t}$ events. The figure is taken from [112].

is defined as the negative vectorial sum over the transverse momenta of all PF particles and transverse energy is the scalar magnitude of the total transverse momentum:

$$\vec{p}_{\mathrm{T}}^{\,miss} = -\sum_{\text{PF particles}} \vec{p}_{\mathrm{T}} \tag{5.7}$$

Since jets are involved in the missing transverse momentum calculation, missing energy is also corrected for jet energy scale [114]. In the following, the absolute value of the missing transverse momentum will be denoted as p_{T}^{miss}.

5.3 EVENT REWEIGHTING

Simulation events are generally tuned to reproduce real data. However, data taking conditions, in particular the PU profile, are not available before data taking starts. Furthermore, the event generation itself is also not perfect. Many details of the hadronization process are still unknown, and state-of-the-art event generators can only compute hard physics processes up to maximally NLO precision, whereas data contains higher orders corrections as well. All these effects can lead to discrepancies between the data and MC simulation.

To correct for some of these discrepancies, event reweighting have been developed. In the following, reweighting applied to the event and the particular physics objects will be described. These include the corrections for mismodeling of the PU, ISR and for a mismatch in efficiency selection for leptons and b-tagged jets in MC and data.

5.3.1 *PU reweighting*

The PU distribution in simulation is randomly sampled from a Poisson distribution with a mean equal to the assumed number of interactions in data. To match correctly the distribution of PU in data, MC simulation needs to be reweighted. The PU distribution in data is calculated using the bunch-by-bunch luminosity, assuming a total proton-proton inelastic cross section of 69.2 mb [63]. For each MC sample, the PU distribution is stored before event selection, and those events which pass through analysis chain are reweighted to match PU profile in data.

5.3.2 *Lepton efficiency and scale factors*

All variables used in the lepton identification, isolation, impact parameters and jet variables have been carefully validated in terms of agreement between data and simulation. Nevertheless, small differences remain. In order to mitigate this effect, correction factors are used to reweight simulated events. These scale factors (SF) are derived for each lepton flavor by measuring efficiencies separately in data and simulation, using the tag-and-probe technique in $Z \to \ell\ell$ events [95, 97].

The Z boson is reconstructed within the invariant mass range of 60-120 GeV from two objects: one lepton passing a tight selection called the tag, and the other lepton passing some looser criterion, than the main one, is called the probe. The tight selection also includes matching to the object selected by the trigger system. The invariant masses of the pair of the tag and passing the selection probe leptons, and the pair of the tag and failing the selection probe leptons are fitted as signal and background. The efficiency is computed by the ratio of the number of signal lepton pairs that both pass tag and probe selection compared to the number of all lepton pairs. This method is applied both to data and simulation. The scale factor is computed as the ratio of the two efficiencies:

$$\mathrm{SF}(p_\mathrm{T}, \eta) = \frac{\varepsilon_\mathrm{data}(p_\mathrm{T}, \eta)}{\varepsilon_\mathrm{MC}(p_\mathrm{T}, \eta)}, \tag{5.8}$$

where $\varepsilon_\mathrm{i}(p_\mathrm{T}, \eta)$ is the efficiency measured in bins of p_T and η. The scale factor is used later to correct the weight of the simulated event. The full simulation correction from the lepton side is given by the product of all scale factors :

$$\mathrm{SF} = \prod_{i \in \mathrm{leptons}} \mathrm{SF}_i(p_\mathrm{T}^i, \eta^i) \tag{5.9}$$

Additional lepton selection efficiencies have been measured for leptons from W decays in simulated $\mathrm{t\bar{t}}$ events as a function of lepton p_T and η to correct the mismatch between measured efficiencies in "FullSim" and "FastSim" simulations. This measurement is cross-checked with the similar measurement in simulated Drell-Yan (DY) events.

5.3.3 *Corrections for b tagging efficiencies*

As a first step towards applying corrections for b tagging efficiencies the distribution of the b tagging discriminator output is reweighted to match the shape for light and heavy flavor jets in MC simulation to the one measured in data. The second component - the b tagging efficiency scale factors for b and light flavor jets - are determined in the

similar way to the lepton efficiencies, using tag and probe technique. The region chosen for this measurement is dominated by the $t\bar{t}$ process, where the electron and muon tight selection is applied. The hardest jets in this control region are supposed to be jets which come from the top quarks decays and are expected to be b-tagged jets. The SF depends on the jet flavor, p_T and η and is applied in the following way. The probability of a given configuration of jets in MC simulation and data is defined as:

$$P(\text{MC}) = \prod_{\text{i=tagged}} \varepsilon_i \prod_{\text{j=not tagged}} (1 - \varepsilon_j)$$

$$P(\text{Data}) = \prod_{\text{i=tagged}} \text{SF}_i \varepsilon_i \prod_{\text{j=not tagged}} (1 - \text{SF}_j \varepsilon_j)$$

where ε is the MC b tagging efficiency and SF is the ratio of b tagging efficiency in data and MC. The event weight is calculated as the ratio of these probabilities in data and simulation.

5.3.4 *ISR corrections*

A study was performed to investigate how well the ISR is modeled in the simulation by evaluating agreement between data and simulation for $t\bar{t}$ production decaying into two lepton final state events. These events are reweighted based on the number of ISR jets in order to make the jet multiplicity agree with data. The same reweighting procedure is applied to SUSY MC events. The reweighting factors vary between 0.92 and 0.51 for the number of ISR jets between 1 and 6. One half of the deviation from unity is considered as the systematic uncertainty on these reweighting factors.

Part I

SEARCH FOR SUPERSYMMETRY IN EVENTS WITH MULTIPLE CHARGED LEPTONS, JETS, AND MISSING TRANSVERSE MOMENTUM

6

SEARCH FOR SUSY IN DATA SET COLLECTED BY CMS EXPERIMENT IN 2016

With its unprecedented collision energies during Run II, the LHC opened once more unexplored territory of SUSY particle masses for our explorations. In this chapter, we discuss the first of the three data analyses performed by the PhD candidate: a search for supersymmetry using collision events with multiple leptons.

There is no 'golden discovery channel' for SUSY. This is because of the unknown mass spectrum of the SUSY particles. Depending on the masses the production rates and the kinematic properties of the final state particles could alter vastly. A systematic search over all possible kinematic phase space was conducted by several research groups with the comprehensive program that was set up by the CMS collaboration. Due to extremely small (< 1 pb) cross sections for SUSY processes it is essential to employ dedicated techniques to discriminate SUSY signal against the background which arises from SM processes. Among several strategies, the usage of leptons appears to be appealing: dominant QCD process and single boson production do not have enough leptons to enter the multilepton event selection. Furthermore, for the processes yielding multiple bosons in the final state, every charged lepton decreases background efficiencies due to the leptonic branching fractions of approximately 11% and 3.4% per lepton flavor for W and Z bosons, respectively [10]. In this thesis extremely low SM background is achieved by the usage of at least three charged leptons in the final state. Unfortunately, high lepton multiplicity also implies low signal output for the SUSY models (see subsection 6.1).

Similar searches as the one presented in this thesis have been carried out by the ATLAS [115] and the CMS [116, 117] Collaborations at $\sqrt{s} = 8$ TeV based on data sets corresponding to 20.3 fb^{-1} and 19.5 fb^{-1} of proton-proton collisions, respectively. All of these searches did not show significant deviations of the observed data from the SM background expectation, which allowed to exclude gluino masses of up to 1 TeV in different models. A first version of the search presented in this thesis has been conducted with the first data set of proton-proton collisions collected by CMS at $\sqrt{s} = 13$ TeV in 2015, totaling an integrated luminosity of 2.3 fb^{-1} [118]. Subsequently, the search has been optimized for a larger data set and has been performed with 35.9 fb^{-1} collected in 2016.

The results presented in this section are published by the CMS collaboration in JHEP. The original paper for which the entire set of results are taken from this thesis work is added here. After the paper, we will explain more in detail each component of the analysis.

EUROPEAN ORGANIZATION FOR NUCLEAR RESEARCH (CERN)

CERN-EP-2017-243
2018/02/21

CMS-SUS-16-041

Search for supersymmetry in events with at least three electrons or muons, jets, and missing transverse momentum in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$

The CMS Collaboration*

Abstract

A search for new physics is carried out in events with at least three electrons or muons in any combination, jets, and missing transverse momentum. Results are based on the sample of proton-proton collision data produced by the LHC at a center-of-mass energy of 13 TeV and collected by the CMS experiment in 2016. The data sample analyzed corresponds to an integrated luminosity of $35.9\,\text{fb}^{-1}$. Events are classified according to the number of b jets, missing transverse momentum, hadronic transverse momentum, and the invariant mass of same-flavor dilepton pairs with opposite charge. No significant excess above the expected standard model background is observed. Exclusion limits at 95% confidence level are computed for four different supersymmetric simplified models with pair production of gluinos or third-generation squarks. In the model with gluino pair production, with subsequent decays into a top quark-antiquark pair and a neutralino, gluinos with masses smaller than 1610 GeV are excluded for a massless lightest supersymmetric particle. In the case of bottom squark pair production, the bottom squark masses are excluded up to 840 GeV for charginos lighter than 200 GeV. For a simplified model of heavy top squark pair production, the $\widetilde{\text{t}}_2$ mass is excluded up to 720, 780, or 710 GeV for models with an exclusive $\widetilde{\text{t}}_2 \to \widetilde{\text{t}}_1\text{H}$ decay, an exclusive $\widetilde{\text{t}}_2 \to \widetilde{\text{t}}_1\text{Z}$ decay, or an equally probable mix of those two decays. In order to provide a simplified version of the analysis for easier interpretation, a small set of aggregate signal regions also has been defined, providing a compromise between simplicity and analysis sensitivity.

Published in the Journal of High Energy Physics as doi:10.1007/JHEP02(2018)067.

*See Appendix A for the list of collaboration members

1 Introduction

Many different theories beyond the standard model (BSM) predict processes leading to events containing multiple electrons and/or muons [1–5]. The background from standard model (SM) processes forging such a final state is small and dominated by multiboson production, which is well understood theoretically [6–20] and well reconstructed experimentally [21–25]. The search in this paper is designed to have broad sensitivity to a variety of BSM models by examining the event yields as a function of several kinematic quantities.

This paper describes the methods and results of a search for new physics in final states with three or more electrons or muons in any combination accompanied by jets and missing transverse momentum. A sample of proton-proton (pp) collision data, corresponding to an integrated luminosity of 35.9 fb^{-1} and collected by the CMS detector at the CERN LHC at a center-of-mass energy of 13 TeV throughout 2016, is used. Results of this analysis are interpreted in the context of supersymmetric (SUSY) models [26–34]. Supersymmetry is an extension of the SM that predicts a SUSY partner for every SM particle by introducing a new symmetry between bosons and fermions. It can potentially provide solutions to questions left open by the SM, such as the hierarchy problem and the nature of dark matter. More specifically, models in which *R*-parity [31] is conserved, whereby SUSY particles are produced only in pairs, can include a dark matter candidate in the form of a stable and undetectable lightest SUSY particle (LSP). In the models considered in this paper, the LSP is assumed to be the lightest neutralino (a mixture of the superpartners of the Higgs and Z bosons, and of the photon).

The reference models for this analysis are simplified model spectra (SMS) [35]. Examples for SUSY processes that can give rise to multilepton final states are shown in Fig. 1. Throughout this paper lepton refers to an electron or a muon. The models under consideration in this analysis feature the pair production of gluinos, $\widetilde{\mathrm{g}}$, or third generation squarks, $\widetilde{\mathrm{b}}_1$ or $\widetilde{\mathrm{t}}_2$, superpartners of gluons and third generation quarks, respectively, for a wide spectrum of possible masses. A typical process predicted by SUSY models consists of gluino pair production with each gluino decaying to a top quark pair, $\mathrm{t\bar{t}}$, and an LSP, $\widetilde{\chi}_1^0$ (Fig. 1, upper left), or to a pair of quarks and a neutralino, $\widetilde{\chi}_2^0$, or chargino, $\widetilde{\chi}_1^{\pm}$. The latter would then decay into a Z or W boson, and an LSP (Fig. 1, upper right). The first model is referred to as T1tttt and the second one as T5qqqqVV throughout this paper. Other models feature bottom squark, $\widetilde{\mathrm{b}}_1$, pair production, with subsequent cascade decays resulting in top quarks, W bosons and LSPs (Fig. 1, lower left) or pair production of the heaviest of the two top squark states, $\widetilde{\mathrm{t}}_2$, with subsequent decays to top quarks, Higgs or Z bosons, and LSPs (Fig. 1, lower right). The latter process allows a challenging scenario to be probed in which the mass difference between the lighter top squark, $\widetilde{\mathrm{t}}_1$, and the neutralino, $\widetilde{\chi}_1^0$, is close to the mass of the top quark [36, 37]. These two models are denoted as T6ttWW and T6ttHZ, respectively. Through the decays of W, Z or Higgs bosons these processes can result in several leptons. In addition to the presence of multiple leptons, these models predict events with multiple jets and missing transverse momentum, largely induced by the undetected LSPs. The SUSY particles that are not directly included in the diagrams are assumed to be too heavy to be accessible at the LHC. Therefore, the only free parameters in these models are the mass of the produced gluinos or squarks, the masses of the possible intermediate particles in the decay chain, like $\widetilde{\chi}_2^0$ or $\widetilde{\chi}_1^{\pm}$, and the mass of the $\widetilde{\chi}_1^0$.

Similar searches have been carried out by the ATLAS and CMS Collaborations using the 13 TeV dataset. With the data sample collected by the ATLAS experiment and corresponding to an integrated luminosity of 36.1 fb^{-1}, gluinos with masses up to 1870 GeV can be excluded [38] assuming the model depicted in Fig. 1 (upper left). A comparable search at the same center-of-mass energy with the CMS detector in 2015, based on a data sample corresponding to an

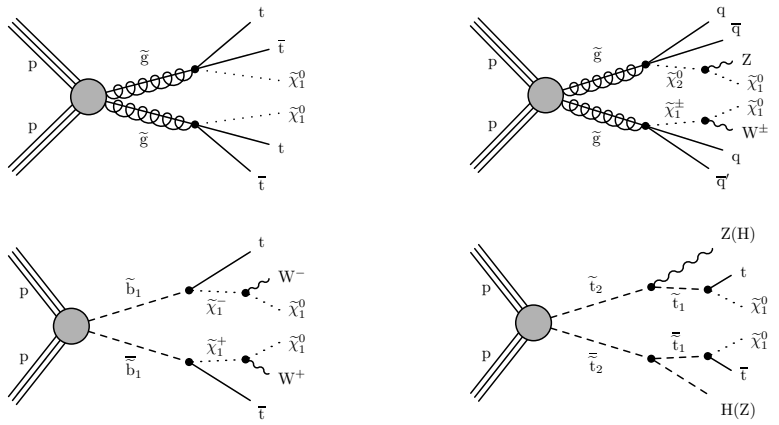


Figure 1: Diagrams for models with gluino pair production leading to four top quarks, T1tttt (upper left), or four quarks and two vector bosons, T5qqqqVV (upper right) in the final state, in both cases accompanied by two LSPs. Models of bottom, T6ttWW, and top squark, T6ttHZ, pair production lead to two top quarks, two LSPs and either two W bosons (lower left) or two neutral bosons as SM Higgs (H) and/or Z bosons (lower right).

integrated luminosity of 2.3 fb^{-1}, excluded gluino masses below 1175 GeV [39]. The current analysis improves upon the one performed with the data collected in 2015 with a more advanced strategy that exploits the transverse mass reconstructed with a lepton and the missing transverse momentum vector. Taking into account that approximately 15 times more data were collected in 2016, a new control region dominated by events from the $\mathrm{t\bar{t}Z}$ process and a new interpretation of the results based on a T6ttHZ model also were added.

2 The CMS detector

The CMS detector features a superconducting solenoid with an internal diameter of 6 m that creates a magnetic field of 3.8 T. Inside the magnet volume are a silicon pixel and strip tracker, an electromagnetic calorimeter (ECAL) made of lead tungstate crystals, and a hadronic calorimeter (HCAL) made of brass and scintillator material, each composed of a barrel and two endcap sections. Forward calorimeters extend the pseudorapidity (η) coverage for the HCAL. In the barrel section of the ECAL, an energy resolution of about 1% is achieved for unconverted or late-converting photons in the tens of GeV energy range. The remaining barrel photons have a resolution of about 1.3% up to $|\eta| = 1$, rising to about 2.5% at $|\eta| = 1.4$. In the endcaps, the resolution of unconverted or late-converting photons is about 2.5%, while the remaining endcap photons have a resolution between 3 and 4% [40]. When combining information from the entire detector, the jet energy resolution amounts typically to 15% at 10 GeV, 8% at 100 GeV, and 4% at 1 TeV, to be compared to about 40%, 12%, and 5% obtained when the ECAL and HCAL calorimeters alone are used. Muons are measured in the range $|\eta| < 2.4$, with detection planes made using three technologies: drift tubes, cathode strip chambers, and resistive plate chambers. Matching muons to tracks measured in the silicon tracker results in a relative transverse momentum resolution for muons with $20 < p_\mathrm{T} < 100$ GeV of 1.3–2.0% in the barrel and better than 6% in the endcaps, The p_T resolution in the barrel is better than 10% for muons with p_T up to 1 TeV [41]. The first level of the CMS trigger system [42], composed of specialized hardware processors, uses information from the calorimeters and muon detectors to select the

most interesting events in a fixed time interval of less than 4 μs. The high-level trigger (HLT) processor farm further decreases the event rate from approximately 100 kHz to around 1 kHz, before the storage of the data. A more detailed description of the CMS detector, together with a definition of the coordinate system used and the relevant kinematic variables, can be found in [43].

3 Event selection criteria and Monte Carlo simulation

Events are reconstructed using the particle flow, PF, algorithm [44], which reconstructs and identifies each individual particle with an optimized combination of information from the various elements of the CMS detector. The objects identified as particles by this algorithm are commonly referred to as PF candidates. Jets are clustered from PF candidates using the anti-k_T algorithm [45, 46] with a distance parameter of 0.4. Only jets with transverse momentum (p_T) larger than 30 GeV falling within $|\eta| < 2.4$ are considered. To avoid double counting, the closest matching jets to leptons are not considered if they are separated from the lepton by less than 0.4 in $\Delta R \equiv \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$. Here $\Delta\eta$ and $\Delta\phi$ are the differences in η and azimuthal angle (ϕ, in radians) between the considered lepton and a given jet. Additional criteria are applied to reject events containing noise and mismeasured jets. Jet energy scale (JES) corrections are applied to correct simulated jets for residual differences with data [47, 48].

The combined secondary vertex algorithm CSVv2 [49, 50] is used to assess the likelihood that a jet originates from a bottom quark. The tagging efficiency for true b flavor jets is typically 70% and the misidentification probabilities are 10% and 1% for c quark and light-flavor jets, respectively. Jets with $p_T > 25$ GeV and within $|\eta| < 2.4$ are considered for b tagging. Another variable related to jets that is used throughout this analysis is the scalar sum of the transverse momenta of all jets, defined as $H_T = \sum_{\text{jets}} p_T$, where jets have $p_T > 30$ GeV. The missing transverse momentum p_T^{miss} is defined as the magnitude of $\vec{p}_T^{\text{miss}}$, the negative vector sum of the transverse momenta all PF candidates reconstructed in an event [51, 52].

Electron candidates are reconstructed using tracking and ECAL information, by combining the clusters of energy deposits in the ECAL with Gaussian sum filter tracks [53]. The electron identification is performed using a multivariate discriminant built with shower shape variables, track-cluster matching variables, and track quality variables. The algorithm is optimized to select electrons from the decay of W and Z bosons with a 90% efficiency while rejecting electron candidates originating from jets. To reject electrons originating from photon conversions inside the detector, electrons are required to have all possible measurements in the innermost tracker layers and to be incompatible with any conversion-like secondary vertices. The identification of the muon is performed using the quality of the matching between the measurements of the tracker and the muon system [41]. The muon identification efficiency is at least 96%, with some variation depending on p_T and η.

The reconstructed vertex with the largest value of summed physics object p_T^2 is taken to be the primary pp interaction vertex. The physics objects are the objects returned by a jet finding algorithm [45, 46] applied to all charged tracks associated with the vertex, plus the corresponding associated missing transverse momentum. Both muon and electron candidates are required to have a transverse impact parameter smaller than 0.5 mm with respect to the primary vertex and a longitudinal impact parameter smaller than 1 mm. In addition, a selection on the three-dimensional impact parameter significance, defined as the value of impact parameter divided by its uncertainty, is applied. This value has to be smaller than 4 for both electrons and muons.

Additional information about the isolation of the lepton is necessary to discriminate between

leptons originating from decays of heavy particles such as W and Z bosons ("prompt" leptons) and those produced in hadron decays or jets misidentified as leptons ("nonprompt" leptons). The lepton isolation criterion is constructed using three different variables.

The relative isolation, I_{rel}, is defined as the ratio of the amount of energy measured in a cone around the lepton to the p_{T} of the lepton, p_{T}^{ℓ}, with a p_{T}^{ℓ}-dependent radius [54]:

$$\Delta R \leq \frac{10\,\text{GeV}}{\min(\max(p_{\text{T}}^{\ell}, 50\,\text{GeV}), 200\,\text{GeV})}. \tag{1}$$

Requiring I_{rel} below a given threshold ensures that the lepton is locally isolated, even in Lorentz-boosted topologies.

The second isolation variable is the ratio of the lepton p_{T} and that of the jet geometrically closest to the lepton: $p_{\text{T}}^{\text{ratio}} = p_{\text{T}}^{\ell}/p_{\text{T}}^{\text{jet}}$. In most cases this is the jet containing the lepton. If no jet is found within a cone defined by $\Delta R < 0.4$, the ratio is set to 1. The use of $p_{\text{T}}^{\text{ratio}}$ provides a way to identify nonprompt low-p_{T} leptons originating from low-p_{T} b jets, which decay with a larger opening angle than the one used in I_{rel}.

The last variable used in the isolation criteria of leptons is $p_{\text{T}}^{\text{rel}}$, defined as the magnitude of the component of the lepton momentum perpendicular to the axis of the closest jet. The jet axis is obtained by subtracting the momentum vector of the lepton from that of the jet. If no matched jet is found around the lepton, the variable is set to 0. This variable allows the recovery of leptons from accidental overlap with jets in Lorentz-boosted topologies. For the calculation of $p_{\text{T}}^{\text{rel}}$ and the previously mentioned $p_{\text{T}}^{\text{ratio}}$, jets with $p_{\text{T}} > 5\,\text{GeV}$ and without any additional identification criteria are considered.

Using those three variables, a lepton is considered isolated if the following condition is fulfilled:

$$I_{\text{rel}} < I_1 \text{ AND } (p_{\text{T}}^{\text{ratio}} > I_2 \text{ OR } p_{\text{T}}^{\text{rel}} > I_3). \tag{2}$$

The values of I_1, I_2, and I_3 depend on the flavor of the lepton; the probability to misidentify a jet as a lepton is higher for electrons than for muons, so tighter isolation values are used for the former. For electrons (muons), the tight selection requirements are $I_1 = 0.12$ (0.16), $I_2 = 0.76$ (0.69), and $I_3 = 7.2$ (6.0) GeV. The isolation requirement for leptons to pass the loose working point of the selection is significantly relaxed, only consisting of $I_{\text{rel}} < 0.4$.

Events used in this analysis are required to pass trigger selection criteria that target dilepton and multilepton events. The following two sets of triggers are used in a logic OR configuration. One set of triggers requires that the two leptons satisfy loose isolation criteria and that the highest-p_{T} (leading) lepton have $p_{\text{T}} > 23$ (17) GeV and the second highest-p_{T} (sub-leading) lepton have $p_{\text{T}} > 12$ (8) GeV for muons (electrons). The second set of triggers places no requirements on the isolation, has a lower p_{T} threshold for both leptons ($p_{\text{T}} > 8\,\text{GeV}$), and requires the H_{T} reconstructed in the trigger to be greater than 300 GeV. With the thresholds on the p_{T} of the leptons and on the H_{T} applied, the efficiency per event is near 100%.

The selection requires the presence of at least three well-identified leptons in the event. The leptons must satisfy p_{T} thresholds that depend on the lepton flavor and the amount of hadronic activity in the event. For events with low hadronic activity ($H_{\text{T}} < 400\,\text{GeV}$), the leading electron (muon) must satisfy $p_{\text{T}} > 25$ (20) GeV and sub-leading electrons (muons) must satisfy $p_{\text{T}} > 15$ (10) GeV. In events with high hadronic activity ($H_{\text{T}} > 400\,\text{GeV}$), the thresholds are relaxed to 15 (10) GeV for the leading electrons (muons). The lowest-p_{T} (trailing) lepton must have

$p_T > 10$ GeV in all cases. Opposite-charge same-flavor lepton pairs are required to have an invariant mass ($m_{\ell\ell}$) greater than 12 GeV to suppress Drell–Yan and quarkonium processes.

In order to estimate the contribution from SM processes with prompt leptons in the signal regions and to calculate the predicted yields from new physics models, Monte Carlo (MC) simulations are used. The MADGRAPH5_aMC@NLO v2.2.2 or v2.3.3 generator [55] was used to simulate events for the $t\bar{t}$, $W\gamma^*$ and tWZ processes, at leading order (LO), and for $t\bar{t}Z$, $t\bar{t}W$, tZq, tHq, tHW, WWZ, WZZ, ZZZ, $t\bar{t}\gamma$, and $Z\gamma^*$ final states, at next-to-leading order (NLO) in perturbative quantum chromodynamics. The NLO POWHEG v2 [56] generator is exploited for the $t\bar{t}H$ [57] and diboson [58, 59] production. The NNPDF3.0LO [60] parton distribution functions (PDFs) are used for the simulated samples generated at LO and the NNPDF3.0NLO [60] PDFs for those generated at NLO. Parton showering and hadronization are simulated using the PYTHIA v8.212 generator [61] with the CUETP8M1 tune [62, 63]. A double-counting of the partons generated with MADGRAPH5_aMC@NLO and those with PYTHIA is removed using the MLM [64] and the FXFX [65] matching schemes, in the LO and NLO samples, respectively. The CMS detector response is modeled using a GEANT4-based model [66]. The simulated samples include additional simultaneous interactions per bunch crossing (pileup), with distributions that are weighted to match the observed data.

Monte Carlo simulation of signal events used for interpretation of the final results is done with the MADGRAPH5_aMC@NLO program at LO precision, allowing for up to two additional partons in the calculation of the matrix elements. The SUSY particle decays, parton showering, and hadronization are simulated with PYTHIA v8.212. The detector response for signal events is simulated using a CMS fast-simulation package [67] that is validated with respect to the GEANT4-based model. All simulated events are processed with the same reconstruction procedure as data. Cross sections for SUSY signal processes, calculated at NLO with next-to-leading-logarithmic (NLL) resummation, were provided by the LHC SUSY Cross Section Working Group [68–73].

4 Search strategy

A baseline selection is applied to the dataset containing events of interest: three or more electrons or muons, at least two jets ($N_{\text{jets}} \geq 2$), $p_T^{\text{miss}} \geq 50$ GeV, and $m_{\ell\ell} \geq 12$ GeV for all opposite-charge, same-flavor lepton pairs. All these requirements are listed in Table 1. Two different regions are defined, based on whether or not an event contains an opposite-charge, same-flavor lepton pair with an invariant mass within the 15 GeV window around the Z boson mass [74]. If such a lepton pair is found the event is categorized as "on-Z", otherwise "off-Z".

Table 1: Summary of all requirements used in baseline selection criteria.

Number of selected leptons	$\geq$3
N_{jets}	$\geq$2
p_T^{miss}, GeV	>50 (70 in low $N_{\text{b jets}}$ and low H_T category)
$m_{\ell\ell}$, GeV	>12

Events are further categorized into signal regions, which are defined according to several event observables: $N_{\text{b jets}}$, H_T, p_T^{miss}, $m_{\ell\ell}$, as well as the transverse mass reconstructed with a lepton and the missing transverse momentum vector,

$$M_T = \sqrt{2p_T^{\ell} p_T^{\text{miss}} \left[1 - \cos\left(\phi_{\ell} - \phi_{\vec{p}_T^{\text{miss}}}\right)\right]}. \quad (3)$$

If the event is categorized as on-Z, the M_T is calculated with the lepton that is not involved in the Z boson mass reconstruction, otherwise the lepton yielding the lowest M_T value (M_T^{min}) is used in the computation of this variable.

The classification of selected events based on the number of b jets creates signal regions with high signal-to-background ratios for events from different signal models. For example, the T1tttt model features several b jets, which would be categorized into signal regions that are almost free of the leptonic WZ background owing to the b jet requirements. Including the 0 b jet signal regions keeps the analysis sensitive to signatures without b jets, such as T5qqqqVV model. Additionally, a categorization in H_T and p_T^{miss} is useful to distinguish between compressed and noncompressed SUSY spectra, i.e. models with small or large mass differences between the SUSY particles in the decay chain.

Table 2 shows the definition of the signal regions (SRs) into which the events passing the baseline selection are subdivided. There are 16 separate off-Z and 16 on-Z SRs. Each category is split, depending on the number of b jets (0, 1 and 2), the value of H_T (greater or lower than 400 GeV), and p_T^{miss} (greater or lower than 150 GeV). These SRs are denoted as SR 1-12. Motivated by the low expected yield of events with high b jet multiplicities, one inclusive SR with $p_T^{miss} < 300$ GeV and $H_T < 600$ GeV has been defined for $\geq$3 b jets (SR 13), and additionally to this three SRs with significant amounts of H_T (>600 GeV, SRs 14, 15) or p_T^{miss} (>300 GeV, SR 16) have been introduced, since various noncompressed SUSY models yield very high values for these variables. These latter three regions are inclusive in the number of b jets. All of the 0 b jet regions, as well as three regions with high H_T and p_T^{miss} values, are further split depending whether M_T is smaller (designated with the letter "a" after the region number) or greater (designated with "b") than 120 GeV, leading to a total of 23 regions for each of the off-Z and on-Z categories. In the on-Z regions with 0 or 1 b jet and $60 < H_T < 400$ GeV, the p_T^{miss} lower bound is raised to 70 GeV to completely suppress the contribution from the Drell–Yan process.

Table 2: Summary of the signal region definitions. The minimum p_T^{miss} requirement is raised from 50 to 70 GeV only for the on-Z SR1 and SR5. Signal regions that are further subdivided at M_T = 120 GeV are indicated with †. The search regions are mirrored for on- and off-Z categories.

N_{jets}	$N_{b\,jets}$	H_T [GeV]	$50(70) \leq p_T^{miss} < 150$ GeV	$150 \leq p_T^{miss} < 300$ GeV	$p_T^{miss} \geq 300$ GeV
$\geq$2	0	60–400	SR1 †	SR2 †	SR16 †
		400–600	SR3 †	SR4 †	
	1	60–400	SR5	SR6	
		400–600	SR7	SR8	
	2	60–400	SR9	SR10	
		400–600	SR11	SR12	
	$\geq$3	60–600	SR13		
	inclusive	$\geq$600	SR14 †	SR15 †	

In order to provide a simplified version of the analysis for easier interpretation, a small set of aggregate signal regions has been defined, providing a compromise between simplicity and analysis sensitivity. The definition of these so-called super signal regions (SSR) is given in Table 3. The additional requirement M_T greater than 120 GeV was added to the SSRs with respect to the relevant SRs.

5 Background estimation

All backgrounds leading to the multilepton final states targeted by this analysis can be subdivided into the categories listed below.

Table 3: Definition of the aggregate super signal regions (SSRs). This simpler classification is proposed for reinterpretations, depending on the presence of a Z boson candidate and the number of b jets, along with additional simultaneous requirements on M_T, p_T^{miss}, and H_T.

	$N_{b\,jets} \leq 2$, $M_T^{min} \geq 120\,GeV$ $H_T \geq 200\,GeV$, $p_T^{miss} \geq 250\,GeV$	$N_{b\,jets} \geq 3$, $M_T^{min} \geq 120\,GeV$ $H_T \geq 60\,GeV$, $p_T^{miss} \geq 50\,GeV$
off-Z	SSR1	SSR2
on-Z	SSR3	SSR4

Nonprompt leptons are leptons from heavy-flavor decays, misidentified hadrons, muons from light-meson decays in flight, or electrons from unidentified photon conversions. In this analysis $t\bar{t}$ events can enter the signal regions if nonprompt leptons are present in addition to the prompt leptons from the W boson decays. Top quark pair production gives the largest contribution for regions with low H_T and p_T^{miss} values, and therefore predominately populates signal regions 1 and 5, with 0 and 1 b jet, respectively. Apart from $t\bar{t}$, Drell–Yan events can enter the baseline selection. However, they are largely suppressed by the $p_T^{miss} > 50\,GeV$ selection, and additional rejection is achieved by increasing the p_T^{miss} requirement to 70 GeV for on-Z regions with low H_T and low p_T^{miss}. Processes that yield only one prompt lepton in addition to nonprompt ones, such as W+jets and various single top quark channels, are effectively suppressed by the three-lepton requirement because of the low probability that two nonprompt leptons satisfy the tight identification and isolation requirements. Albeit small, this contribution is nevertheless accounted for in our method to estimate the background due to nonprompt leptons (see below).

Diboson production can yield multilepton final states with up to three prompt leptons (WZ or Wγ^*) and up to four prompt leptons (ZZ or Zγ^*), rendering irreducible backgrounds for this analysis. For simplicity, in the following we refer to these backgrounds as WZ and ZZ, respectively. The WZ production has a sizable contribution in the on-Z events, especially in the SRs without b jets. The yields of these backgrounds in the various SRs are estimated by means of MC simulation, with the normalization factors derived from control regions in data.

Other rare SM processes that can yield three or more leptons are $t\bar{t}W$, $t\bar{t}Z$, and triboson production. We also include the contribution from the SM Higgs boson produced in association with a vector boson or a pair of top quarks in this category of backgrounds, as well as processes that produce additional leptons from internal conversions, which are events that contain a virtual photon that decays to leptons. The internal conversion background components, X+γ, are strongly suppressed by the $p_T^{miss} > 50\,GeV$ and $N_{jets} \geq 2$ requirements. The background events containing top quark(s) in association with a W, Z or Higgs boson or another pair of top quarks are denoted as $t\bar{t}X$, except for $t\bar{t}Z$ which is separately delineated. For the estimation of the latter process, the same strategy as for the WZ is used. All other processes are grouped into one category that is denoted as rare SM processes. The contribution from these processes as well as $t\bar{t}X$ are estimated from MC simulation.

The background contribution from nonprompt leptons is estimated using the tight-to-loose ratio method [54]. In this method, the yield is estimated in an application region that is similar to the signal region but which contains at least one lepton that fails the tight identification and isolation requirements but satisfies the loose requirements. The events in this region are weighted by $f/(1-f)$, where the tight-to-loose ratio f is the probability that a loosely identified lepton also satisfies the full set of requirements. This ratio is measured as a function of lepton p_T and η in a control sample of multijet events that is enriched in nonprompt leptons (measurement region). In this region, we require exactly one lepton, satisfying the loose object selection, and

one recoiling jet with $\Delta R(\text{jet}, \ell) > 1.0$ and $p_{\mathrm{T}} > 30\,\text{GeV}$ in the event. To suppress processes that can contribute prompt leptons from a W or Z boson decay, such as W(+jets), DY or $\mathrm{t\bar{t}}$, we additionally require both $p_{\mathrm{T}}^{\mathrm{miss}}$ and M_{T} to be below 20 GeV. The remaining contribution from these processes within the measurement region is estimated from MC simulation and subsequently subtracted from the data.

In order to reduce the dependence of the tight-to-loose ratio on the flavor composition of the jets from which the nonprompt leptons originate, this ratio is parameterized as a function of a variable that correlates more strongly with the mother parton p_{T} than with the lepton p_{T}. This variable is calculated by correcting the lepton p_{T} as a function of the energy in the isolation cone around it. This definition leaves the p_{T} of the leptons satisfying the tight isolation criteria unchanged and modifies the p_{T} of those failing these criteria so that it is a better proxy for the mother parton p_{T} and results in a smaller variation as a function of the mother parton p_{T}. The flavor dependence, which is much more important for the case of electrons, is further reduced by adjusting the loose electron selection to obtain similar f values for nonprompt electrons that originate from light- or heavy-flavor jets. As a result, the tight-to-loose ratio measured in a multijet sample leads to a good description of nonprompt background originating from $\mathrm{t\bar{t}}$ events, which in most of the SR are dominant in this category of background.

The tight-to-loose ratio method for estimating the nonprompt background is validated both in a closure test in simulation and in a data control region orthogonal to the baseline selection with minimal signal contamination. This region is defined by the requirement of three leptons that satisfy the nominal identification, isolation, and p_{T} selection, one or two jets, $30 < p_{\mathrm{T}}^{\mathrm{miss}} < 50\,\text{GeV}$, and no dilepton pair with an invariant mass compatible with a Z boson. With these selection criteria a purity in $\mathrm{t\bar{t}}$ of 80% can be achieved. We find an agreement of the order of 20–30% between the predicted and observed yields in this control region.

The WZ process is one of the main backgrounds in the regions with 0 b jets, while $\mathrm{t\bar{t}Z}$ gives a significant contribution in categories enriched in b jets. As mentioned earlier, the contribution of these backgrounds is estimated from simulation, but their normalizations are obtained from a simultaneous fit using two control regions, designed so that each is highly enriched in one of the processes. The WZ control region is defined by the requirement of three leptons satisfying the nominal identification and isolation selections. Two leptons have to form an opposite charge, same flavor pair with $|m_{\ell\ell} - m_{\mathrm{Z}}| < 15\,\text{GeV}$, the number of jets and b jets has to be $\leq$1 and 0, respectively. The $p_{\mathrm{T}}^{\mathrm{miss}}$ has to be in the range $30 < p_{\mathrm{T}}^{\mathrm{miss}} < 100\,\text{GeV}$, and M_{T} is required to be at least 50 GeV to suppress contamination from the Drell–Yan process. The purity of the WZ control region is 80%. The orthogonal control region for $\mathrm{t\bar{t}Z}$ is defined similarly to that for WZ, except for a requirement on the number of jets: three leptons satisfying the nominal identification and isolation selection are to be found, two of them forming an opposite charge, same flavor pair with $|m_{\ell\ell} - m_{\mathrm{Z}}| < 15\,\text{GeV}$, at least 3 jets, and $30 < p_{\mathrm{T}}^{\mathrm{miss}} < 50\,\text{GeV}$. Events are classified by the number of b jets, and three bins are formed for the $\mathrm{t\bar{t}Z}$ CR: the 0 b jet category, where the background is dominated by the WZ and $\mathrm{t\bar{t}}$ processes, and the 1 and $\geq$2 b jet categories, enriched in $\mathrm{t\bar{t}Z}$. The overall purity of the $\mathrm{t\bar{t}Z}$ process is 20%, increasing to 50% in the bins with at least one b jet. These three bins, together with the WZ control region are used in a simultaneous fit to obtain the scale factors for the normalization of the simulated samples. In the fit to data, the normalization and relative population across all four bins of all the components are allowed to vary according to experimental and theoretical uncertainties. For the WZ process the obtained scale factor is compatible with unity, 1.01 ± 0.07, and no correction is applied to the simulation, while for the $\mathrm{t\bar{t}Z}$ it is found to be 1.14 ± 0.28. Therefore the yields from the MC $\mathrm{t\bar{t}Z}$ sample obtained in the baseline region are scaled by a factor of 1.14.

6 Systematic uncertainties

The uncertainties in the expected SM backgrounds and signal yields are categorized as experimental, such as those related to the JES or the b tagging efficiency description in the simulation; theoretical, such as the uncertainties in the considered cross sections; statistical, related to the observed yield in control regions in data; and as uncertainties in the background estimation methods relying on control regions in data. These uncertainties and their effect on the predicted yields are described below and summarized in Table 4.

One of the major experimental sources of uncertainty is the knowledge of the JES. This uncertainty affects all simulated background and signal events. For the data set used in this analysis, the uncertainties in the jet energy scale vary from 1% to 8%, depending on the transverse momentum and pseudorapidity of the jet. The impact of these uncertainties is assessed by shifting the jet energy correction factors for each jet up and down by one standard deviation and recalculating all kinematic quantities. The systematic uncertainties related to JES corrections are also propagated to the $p_{\mathrm{T}}^{\mathrm{miss}}$ calculation. The propagation of the variation of the JES results in a variation of 1–10% in the predicted event yields in the various signal regions of this analysis.

A similar approach is used for the uncertainties associated with the corrections for the b tagging efficiencies for light, charm and bottom flavor jets, which are parameterized as a function of p_{T} and η. The variation of the scale factor correcting for the differences between data and simulation is at a maximum of the order of 10% per jet, and leads to an overall effect in the range of 1–10% depending on the signal region and on the topology of the event. The inaccuracy of the inelastic cross section value that affects the pile up rate gives up to a 5% effect. The sources of uncertainties explained here were also studied for the signal samples, and their impact on the predicted signal yields in every search region has been estimated following the same procedures.

Lepton identification and isolation scale factors have been measured as a function of lepton

Table 4: The effect of the systematic uncertainties on the event yields of the backgrounds and signal processes.

Source	Effect on the backgrounds [%]	Effect on signal [%]
Integrated luminosity	2.5	2.5
JES	1–8	1–10
b tag efficiency	1–8	1–10
Pileup	1–5	1–5
Lepton efficiencies	9	15
HLT efficiencies	3	3
Nonprompt application region statistics	10–100	—
Nonprompt extrapolation	30	—
WZ control region normalization	10	—
$\mathrm{t\bar{t}Z}$ control region normalization	25	—
Limited size of simulated samples	1–100	10–100
ISR modeling	—	1–10
Modeling of unclustered energy	—	1–20
Ren., fact. scales, cross section ($\mathrm{t\bar{t}W}$, $\mathrm{t\bar{t}H}$)	11–13	—
Ren., fact. scales, acceptance ($\mathrm{t\bar{t}W}$, $\mathrm{t\bar{t}Z}$, $\mathrm{t\bar{t}H}$, signal)	3–18	3–18
PDFs ($\mathrm{t\bar{t}W}$, $\mathrm{t\bar{t}Z}$, $\mathrm{t\bar{t}H}$)	2–3	—
Other rare backgrounds	50	—

p_T and η. They are applied to correct for residual differences in lepton selection efficiencies between data and simulation. The corresponding uncertainties are estimated to be about 3% per lepton for both flavors, and additionally 2% per lepton is assigned to the signal leptons due to the detector fast simulation. Assuming 100% correlation between the uncertainties on the corrections for the different leptons, a flat uncertainty of 9% is taken into account for the background, while 15% is considered for the signal. The uncertainty related to the HLT trigger efficiency is evaluated to amount to 3%.

For the nonprompt and misidentified lepton background, several systematic uncertainties are considered. The statistical uncertainty from the application region, which is used to estimate this background contribution, ranges from 10 to 100%. The regions where these uncertainties are large are generally regions where the overall contribution from this background is small. The uncertainty arising from the electroweak background subtraction in the measurement region for the tight-to-loose ratio is propagated from the uncertainty on the scale factor obtained from the fit to the control regions. In the case where no events are observed in the application region, an upper limit of the background expectation is used as determined from the upper limit at 68% confidence level (CL) multiplied by the most likely tight-to-loose ratio value.

The systematic uncertainty related to the extrapolation from the control regions to the signal regions for the nonprompt lepton background is estimated to be 30%. This value has been extracted from closure tests performed by applying the method described in Section 5 to simulated samples containing nonprompt leptons. From the simultaneous fit in the control regions, the uncertainty in the normalization of the WZ process is estimated to be 10%, while a value of 25% is found for $t\bar{t}Z$ background.

The limited size of the generated MC samples represents an additional source of uncertainty. For the backgrounds that are estimated from simulation, such as $t\bar{t}W$, $t\bar{t}Z$ and $t\bar{t}H$, as well as for all the signal processes, this statistical uncertainty is computed from the number of MC events entering the signal regions and varies widely across the SRs.

For signal efficiency calculations additional uncertainties in the description of the initial-state radiation (ISR) are taken into account. The modeling of ISR by the version of the MADGRAPH5_aMC@NLO generator used for signal events was compared against a data sample of $t\bar{t}$ events in the dilepton final state. The corresponding corrections range from 0.51 to 0.92, depending on the jet multiplicity. These corrections are then applied on simulated SUSY events based on the number of ISR jets to improve upon the MADGRAPH5_aMC@NLO modeling of the multiplicity of additional jets from ISR. Half the magnitude of these ISR corrections is assigned as an additional systematic uncertainty, which can be as large as 10%.

The uncertainty in potential differences between the modeling of p_T^{miss} in data and the fast simulation arising from unclustered energy in the CMS detector is evaluated by comparing the reconstructed p_T^{miss} with the p_T^{miss} obtained using generator-level information. This uncertainty ranges up to 20%.

Theoretical uncertainties include the uncertainty in the renormalization (μ_R) and factorization (μ_F) scales, and in the knowledge of the PDFs. These uncertainties are evaluated for several processes, namely $t\bar{t}W$, $t\bar{t}Z$, and $t\bar{t}H$, which are dominant backgrounds in several signal regions. Both the changes in the acceptance and cross sections related to these effects are taken into account and propagated to the final uncertainties.

For the study of the renormalization and factorization uncertainties, variations up and down by a factor of two with respect to the nominal values of μ_F and μ_R are evaluated. The maximum difference in the yields with respect to the nominal case is observed when both scales are varied

up and down simultaneously. The effect on the overall cross section is found to be ∼13% for $\mathrm{t\bar{t}W}$ and ∼11% for $\mathrm{t\bar{t}H}$ backgrounds. The effect of the variations of μ_F and μ_R on the acceptance is taken as additional, uncorrelated uncertainty on the acceptance corresponding to different signal regions. This effect is found to vary between 3% and 18% depending on the SR and the process.

The uncertainty related to the PDFs is estimated from the 100 NNPDF 3.0 replicas, computing the deviation with respect to the nominal yield for each of them in every signal region (the cross section and acceptance effect are considered together) [60]. The root-mean-square of the variations is taken as the value of the systematic uncertainty. Since no significant differences between signal regions have been found, a flat uncertainty of 3% (2%) is considered for $\mathrm{t\bar{t}W}$ ($\mathrm{t\bar{t}Z}$ and $\mathrm{t\bar{t}H}$) backgrounds. This value also includes the effect of the strong coupling constant variation, $\alpha_S(M_Z)$, which is added in quadrature. An extra, conservative, flat uncertainty of 50% is assigned to the yield of the remaining rare processes, which are not well measured.

7 Results

Comparisons between data and the predicted background of the distributions of the four event observables used for signal region categorization, namely H_T, p_T^{miss}, M_T and $N_{b\,jets}$, as well as the lepton p_T spectra, the lepton flavor composition, and the event jet multiplicity are shown in Fig. 2 (Fig. 3) for events satisfying the selection criteria of the off-Z (on-Z). Figure 4 graphically presents a summary of the predicted background and observed event yields in the individual SR bins. The same information is also presented in Tables 5 and 6 for the off-Z and on-Z regions, respectively. Table 7 represents the yields in the SSRs.

The number of events observed in data is found to be consistent with the predicted background yields in all 46 SRs. The results of the search are interpreted by setting limits on superpartner masses using simplified models. For each mass point, the observations, background predictions, and expected signal yields from all on-Z and off-Z search regions are combined to extract the minimum cross section that can be excluded at a 95% CL using the CL_s method [75–77], in which asymptotic approximations for the distribution of the test-statistic, which is a ratio of profiled likelihoods, are used [78]. Log-normal nuisance parameters are used to describe the uncertainties listed in Section 6.

The limits are shown in Fig. 5 for the T1tttt model (left) and for the T5qqqqVV model (right). In the T5qqqqVV model each gluino decays to a pair of light quarks and a neutralino ($\widetilde{\chi}_2^0$) or chargino ($\widetilde{\chi}_1^\pm$), followed by the decay of that neutralino or chargino to a W or Z boson, respectively, and an LSP (Fig. 1, top right). The probability for the decay to proceed via the $\widetilde{\chi}_1^+$, $\widetilde{\chi}_1^-$, or $\widetilde{\chi}_2^0$ is taken to be 1/3 for each case. In this scenario, the second neutralino $\widetilde{\chi}_2^0$ and chargino are assumed to be mass-degenerate, with masses equal to $0.5(m_{\tilde{g}} + m_{\widetilde{\chi}_1^0})$.

The limits on the bottom squark pair production cross section are shown in Fig. 6. In this model, the mass of the LSP is set to 50 GeV. Finally, the limits on the $\widetilde{t}_2$ pair production cross section are shown in Fig. 7. In this scenario, the mass difference between the $\widetilde{t}_1$ and the LSP is set to 175 GeV, the $\widetilde{t}_1$ decays via a top quark to LSP, and the $\widetilde{t}_2$ decays via a Z or Higgs boson to $\widetilde{t}_1$. We consider the reference values $\mathcal{B}(\widetilde{t}_2 \to \widetilde{t}_1 Z) = 0$, 50, and 100%; the sensitivity is diminished for the $\widetilde{t}_1 H$ final state because of the additional branching factors for Higgs cascade decays to electrons or muons via gauge bosons or tau leptons.

Search regions providing the best sensitivity to new physics scenarios depend on the considered models and their parameters. In the non-compressed scenario of the T1tttt model, the most

sensitive region is off-Z SR16b (high p_T^{miss} and M_T region). When considering the compressed scenario, the contribution from SR16b region remains the largest, up to the most compressed cases where the SR12 off-Z region (2 b jets, medium p_T^{miss} and high H_T) starts to contribute significantly. For the T5qqqqVV model in the non-compressed scenario, the most sensitive regions are on-Z SR16b and SR15b (high and medium p_T^{miss}, high H_T and high M_T values). When moving towards more compressed scenarios, the most significant contributions come from the SR16b and SR15b on-Z regions, until reaching the compressed scenario where the most sensitive region is SR4b (medium p_T^{miss}, high H_T and high M_T). The exclusion limit for T6ttWW model is dominated by both off-Z SR16 regions (high p_T^{miss} region). For the T6ttHZ model with $\mathcal{B}(\tilde{t}_2 \to \tilde{t}_1 Z)$ = 0%, the limits in the non-compressed scenario are driven by the off-Z SR15a (high H_T, medium p_T^{miss}, low M_T), while for compressed case by off-Z SR13 (high $N_{b\,jets}$, low and medium H_T and p_T^{miss}). For $\mathcal{B}(\tilde{t}_2 \to \tilde{t}_1 Z)$ = 50% in the non-compressed scenario, the on-Z SR16b region dominates the the exclusion limit, while in the compressed scenario the on-Z SR13 (high $N_{b\,jets}$) and SR15b (high H_T, medium p_T^{miss}, high M_T) give the highest contribution. Finally, for $\mathcal{B}(\tilde{t}_2 \to \tilde{t}_1 Z)$ = 100% the on-Z SR16b plays the leading role in both compressed and non-compressed scenarios.

Table 5: Expected and observed yields in the off-Z search regions. The first uncertainty states the statistical uncertainty, while the second represents the systematic uncertainty.

$N_{b\,jets}$	H_T [GeV]	p_T^{miss} [GeV]	M_T [GeV]	Expected [events]	Observed [events]	SR
0	60–400	50–150	<120	$206 \pm 6 \pm 35$	201	SR1a
			≥120	$1.4 \pm 0.5 \pm 0.2$	3	SR1b
		150–300	<120	$25.9 \pm 2.1 \pm 4.3$	24	SR2a
			≥120	$0.84 \pm 0.34 \pm 0.12$	0	SR2b
	400-600	50–150	<120	$15.6 \pm 1.6 \pm 2.1$	21	SR3a
			≥120	$0.19 \pm 0.09 \pm 0.02$	0	SR3b
		150–300	<120	$6.0 \pm 0.8 \pm 0.7$	5	SR4a
			≥120	$0.19 \pm 0.09 \pm 0.04$	0	SR4b
1	60–400	50–150	Inclusive	$202 \pm 6 \pm 44$	191	SR5
		150–300		$25.6 \pm 1.9 \pm 4.6$	25	SR6
	400-600	50–150		$15.4 \pm 1.3 \pm 2.2$	21	SR7
		150–300		$7.3 \pm 1 \pm 1.1$	7	SR8
2	60–400	50–150	Inclusive	$47.7 \pm 2.8 \pm 7.6$	51	SR9
		150–300		$5.3 \pm 0.5 \pm 0.6$	5	SR10
	400-600	50–150		$5.8 \pm 0.7 \pm 0.8$	9	SR11
		150–300		$2.9 \pm 0.5 \pm 0.4$	2	SR12
≥3	60–600	50–300	Inclusive	$3.9 \pm 0.7 \pm 0.6$	6	SR13
Inclusive	≥600	50–150	<120	$14.4 \pm 1.2 \pm 1.6$	20	SR14a
			≥120	$0.28 \pm 0.14 \pm 0.04$	0	SR14b
		150–300	<120	$12.1 \pm 1.4 \pm 1.6$	10	SR15a
			≥120	$0.40 \pm 0.12 \pm 0.05$	0	SR15b
	≥60	≥300	<120	$12.1 \pm 1.5 \pm 1.9$	7	SR16a
			≥120	$0.70 \pm 0.25 \pm 0.11$	0	SR16b

Table 6: Expected and observed yields in the on-Z search regions. The first uncertainty states the statistical uncertainty, while the second represents the systematic uncertainty.

$N_{\text{b jets}}$	H_T [GeV]	p_T^{miss} [GeV]	M_T [GeV]	Expected [events]	Observed [events]	SR
0	60-400	70-150	<120	$266 \pm 5 \pm 39$	241	SR1a
			≥120	$30 \pm 2 \pm 4$	33	SR1b
		150–300	<120	$53.8 \pm 2.2 \pm 8$	61	SR2a
			≥120	$5.7 \pm 0.8 \pm 0.7$	9	SR2b
	400-600	50–150	<120	$44.6 \pm 1.9 \pm 6.5$	52	SR3a
			≥120	$5.1 \pm 0.6 \pm 0.7$	6	SR3b
		150–300	<120	$16.6 \pm 1.3 \pm 2.5$	17	SR4a
			≥120	$1.43 \pm 0.33 \pm 0.2$	1	SR4b
1	60–400	70-150	Inclusive	$116 \pm 4 \pm 15$	115	SR5
		150–300		$21.7 \pm 1.2 \pm 2.8$	19	SR6
	400-600	50–150		$25.2 \pm 1.2 \pm 3.6$	25	SR7
		150–300		$7.5 \pm 0.8 \pm 1$	9	SR8
2	60–400	50–150	Inclusive	$47 \pm 1.6 \pm 7.4$	64	SR9
		150-300		$7.2 \pm 0.8 \pm 1.2$	6	SR10
	400-600	50–150		$11.7 \pm 1 \pm 2.1$	12	SR11
		150–300		$2.6 \pm 0.4 \pm 0.4$	6	SR12
≥3	60–600	50–300	Inclusive	$4.7 \pm 0.5 \pm 0.9$	5	SR13
Inclusive	≥600	50–150	<120	$33 \pm 2 \pm 4$	42	SR14a
			≥120	$4.6 \pm 0.6 \pm 0.6$	6	SR14b
		150–300	<120	$15.8 \pm 1.2 \pm 2$	13	SR15a
			≥120	$1.9 \pm 0.3 \pm 0.2$	4	SR15b
	≥60	≥300	<120	$19.1 \pm 1.1 \pm 2.8$	23	SR16a
			≥120	$2.28 \pm 0.35 \pm 0.26$	5	SR16b

Table 7: Expected and observed yields in the super signal regions. The background events containing top quark(s) in association with a W, Z or Higgs boson, except $t\bar{t}Z$, or another pair of top quarks are denoted as $t\bar{t}X$. The first uncertainty states the statistical uncertainty, while the second represents the systematic uncertainty.

	SSR1	SSR2	SSR3	SSR4
Nonprompt	$0.63 \pm 0.38 \pm 0.19$	$0.00 \pm 0.00^{+0.3}_{-0.0}$	$0.46 \pm 0.37 \pm 0.14$	$0.21^{+0.23}_{-0.21} \pm 0.06$
$t\bar{t}Z$	$0.14 \pm 0.06 \pm 0.03$	$0.05 \pm 0.03 \pm 0.01$	$1.27 \pm 0.18 \pm 0.31$	$0.54 \pm 0.10 \pm 0.13$
$t\bar{t}X$	$0.23 \pm 0.04 \pm 0.05$	$0.11 \pm 0.04 \pm 0.02$	$0.50 \pm 0.07 \pm 0.08$	$0.17 \pm 0.03 \pm 0.02$
WZ	$0.01 \pm 0.01 \pm 0.01$	$0.01 \pm 0.01 \pm 0.01$	$1.03 \pm 0.28 \pm 0.21$	$0.01 \pm 0.01 \pm 0.01$
Rare	$0.12 \pm 0.06 \pm 0.05$	$0.01 \pm 0.01 \pm 0.01$	$0.40 \pm 0.09 \pm 0.14$	$0.01 \pm 0.01 \pm 0.01$
Total	$1.1 \pm 0.4 \pm 0.2$	$0.18 \pm 0.05^{+0.3}_{-0.02}$	$3.7 \pm 0.5 \pm 0.4$	$0.94^{+0.26}_{-0.23} \pm 0.15$
Observed	0	0	6	2

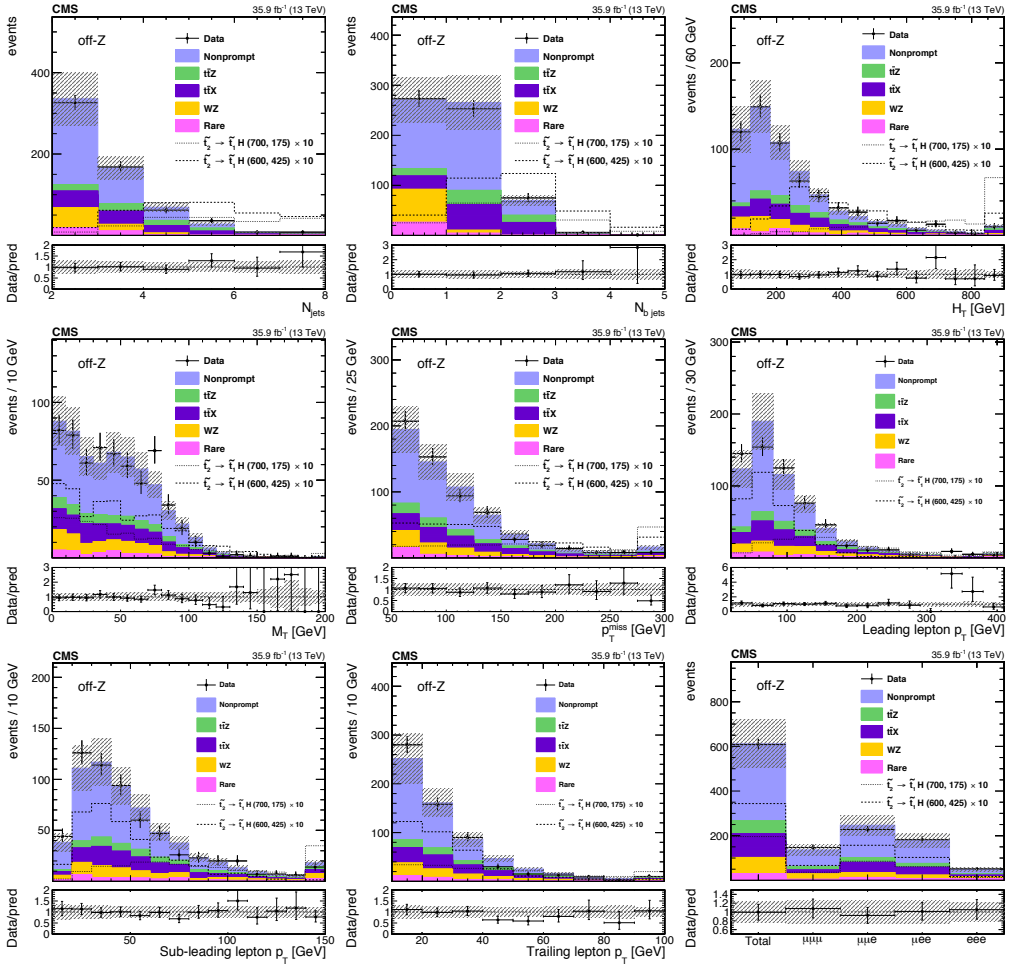


Figure 2: Background prediction and the observed event yields in the key observables for the off-Z baseline selection: the number of jets and b jets, H_T, M_T, p_T^{miss}, the lepton p_T spectra and the event yields by flavor category are shown. The background events containing top quark(s) in association with a W, Z or Higgs boson, except $t\bar{t}Z$, or another pair of top quarks are denoted as $t\bar{t}X$. The last bin includes the overflow events, and the hatched area represents the statistical and combined systematic uncertainties in the prediction. The lower panels show the ratio of the observed and predicted yields in each bin. For illustration the yields, multiplied by a factor 10, for two signal mass points in the T6ttHZ model, where the $\mathcal{B}(\tilde{t}_2 \to \tilde{t}_1 H) = 100\%$, are displayed for non-compressed ($m(\tilde{t}_2) = 700\,GeV$ and $m(\tilde{t}_1) = 175\,GeV$) and compressed ($m(\tilde{t}_2) = 600\,GeV$ and $m(\tilde{t}_1) = 425\,GeV$) scenarios.

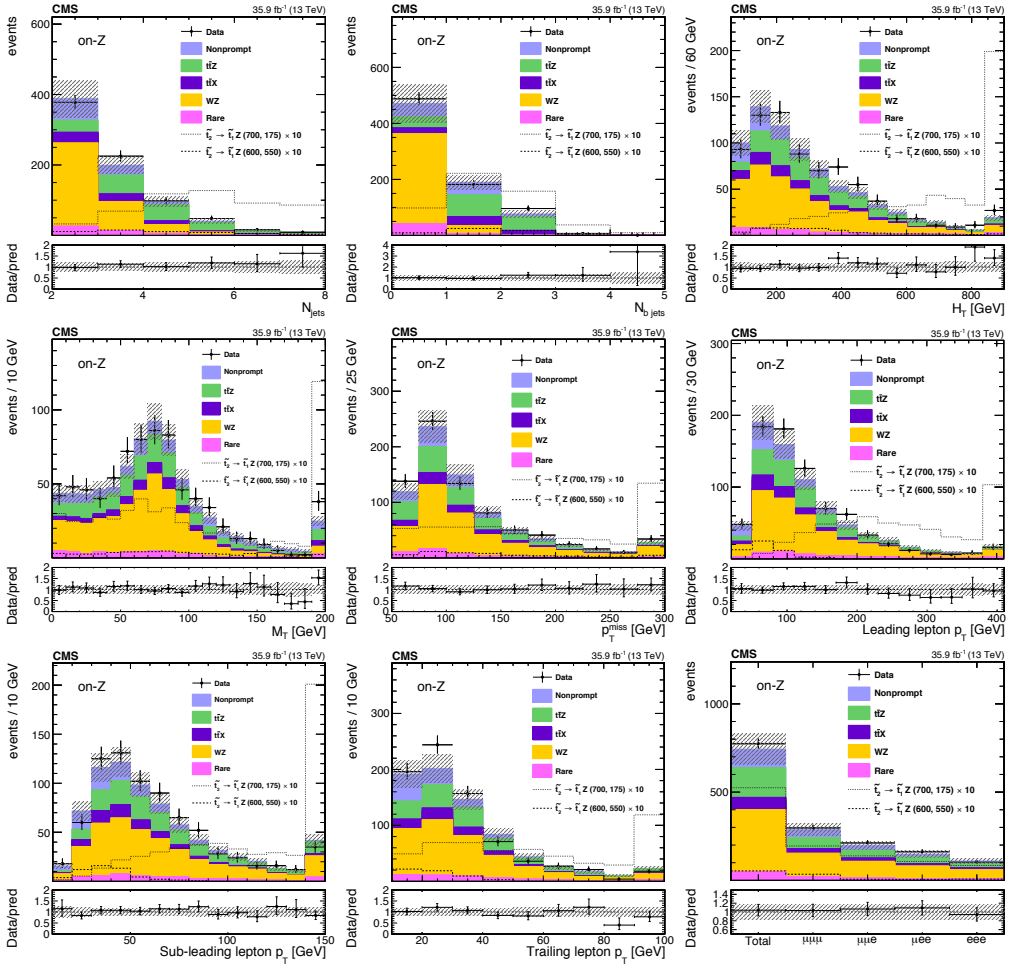


Figure 3: Background prediction and the observed event yields in the key observables of the on-Z baseline selection: the number of jets and b jets, H_T, M_T, p_T^{miss}, the lepton p_T spectra and the event yields by flavor category are shown. The background events containing top quark(s) in association with a W, Z or Higgs boson, except $t\bar{t}Z$, or another pair of top quarks are denoted as $t\bar{t}X$. The last bin includes the overflow events, and the hatched area represents the combined statistical and systematic uncertainties in the prediction. The lower panels show the ratio of the observed and predicted yields in each bin. For illustration the yields, multiplied by a factor 10, for two signal mass points in the T6ttHZ model, where the $\mathcal{B}(\tilde{t}_2 \to \tilde{t}_1 Z) = 100\%$, are displayed for non-compressed ($m(\tilde{t}_2) = 700\,GeV$ and $m(\tilde{t}_1) = 175\,GeV$) and compressed ($m(\tilde{t}_2) = 600\,GeV$ and $m(\tilde{t}_1) = 550\,GeV$) scenarios.

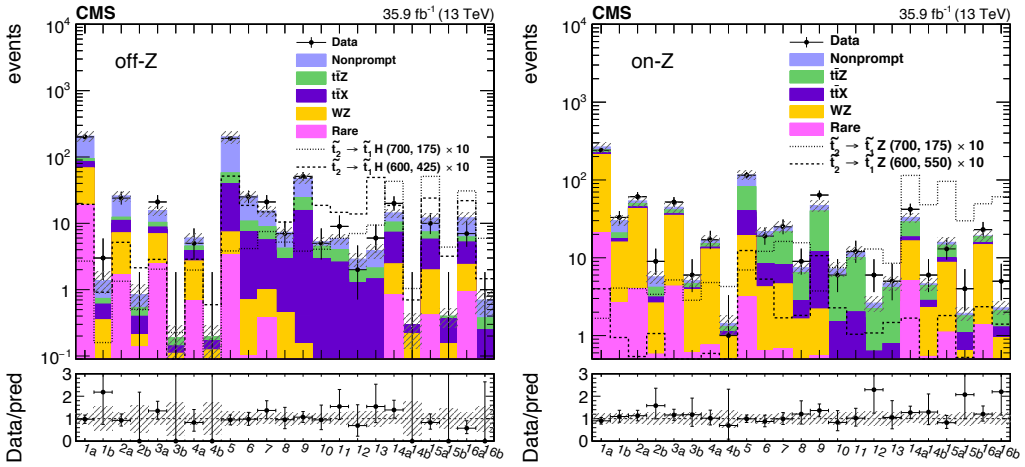


Figure 4: Background prediction and observed event yields in the 23 off-Z (left) and the 23 on-Z (right) signal regions. The background events containing top quark(s) in association with a W, Z or Higgs boson, except $t\bar{t}Z$, or another pair of top quarks are denoted as $t\bar{t}X$. The hatched area represents the statistical and systematic uncertainties on the prediction. The lower panels show the ratio of the observed and predicted yields in each bin. For illustration the yields, multiplied by a factor 10, for $\tilde{t}_2 \to \tilde{t}_1 H$ (left) and $\tilde{t}_2 \to \tilde{t}_1 Z$ (right) decays are displayed for two signal mass points in the T6ttHZ model to represent compressed and non-compressed scenarios.

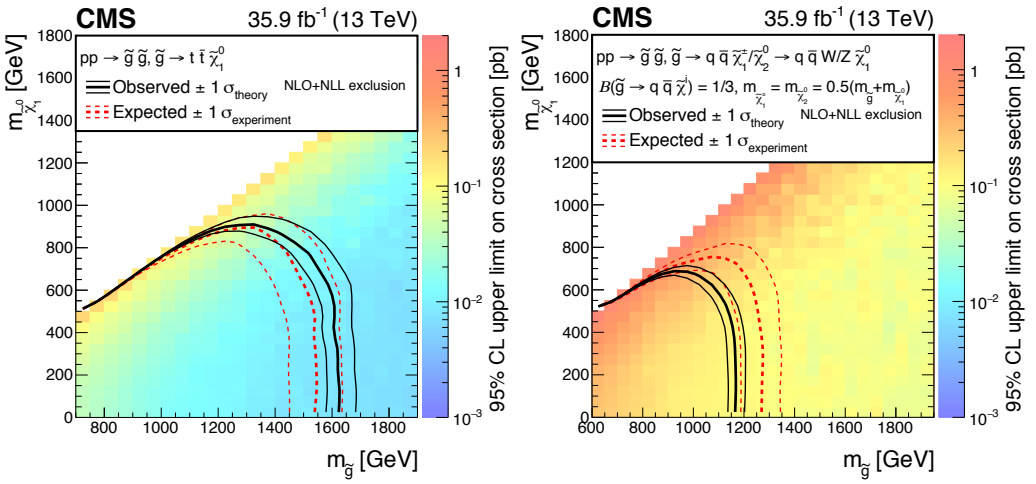


Figure 5: Cross section upper limits at 95% CL in the $m_{\tilde{\chi}_1^0}$ versus $m_{\tilde{g}}$ plane for T1tttt (left) and T5qqqqVV (right) simplified models. For the latter model the branching fraction of gluino decay to neutralino or chargino is equal to 1/3 and $m_{\tilde{\chi}_1^\pm} = m_{\tilde{\chi}_2^0} = 0.5(m_{\tilde{g}} + m_{\tilde{\chi}_1^0})$. The excluded regions are to the left and below the observed and expected limit curves. The color scale indicates the excluded cross section at a given point in the mass plane.

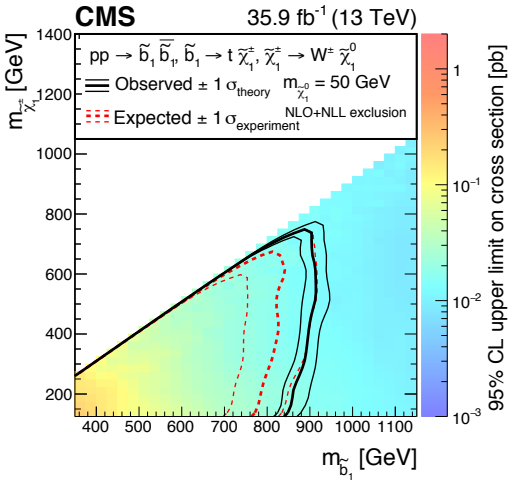


Figure 6: Cross section upper limits at 95% CL in the $m_{\tilde{\chi}_1^\pm}$ versus $m_{\tilde{b}_1}$ plane for T6ttWW simplified model. The mass of the neutralino is set to 50 GeV. The descriptions of the excluded regions and color scale are the same as in Fig. 5.

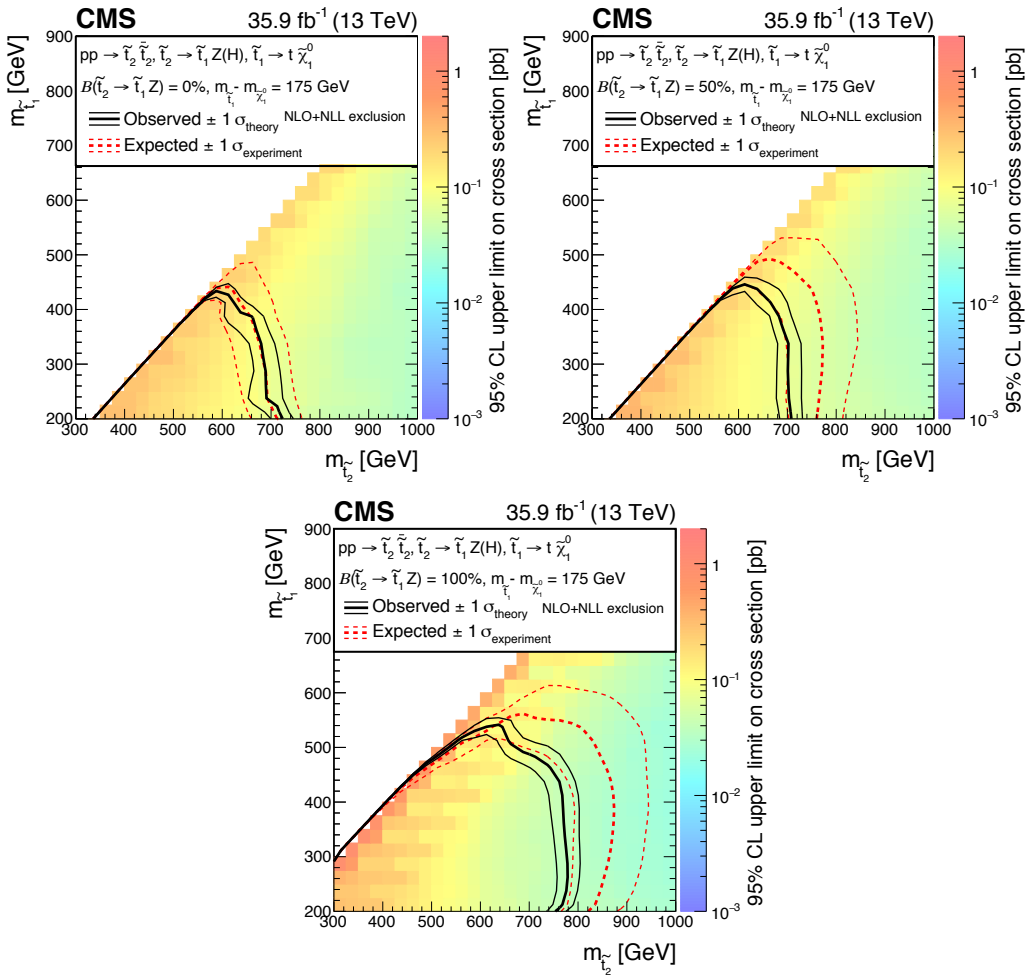


Figure 7: Cross section upper limits at 95% CL in the $m_{\tilde{t}_1}$ versus $m_{\tilde{t}_2}$ plane for T6ttHZ simplified model. Different branching fractions of the decay $\tilde{t}_2 \to \tilde{t}_1 Z$ are considered: 0% (top left), 50% (top right), and 100% (bottom). The mass difference between the lighter top squark ($\tilde{t}_1$) and a neutralino is close to the mass of the top quark. The descriptions of the excluded regions and color scale are the same as in Fig. 5.

8 Conclusions

A search for physics beyond the standard model in final states with at least three electrons or muons in any combination, jets, and missing transverse momentum has been presented using data collected by the CMS detector in 2016 at $\sqrt{s} = 13\,\text{TeV}$, corresponding to an integrated luminosity of $35.9\,\text{fb}^{-1}$. The analysis makes use of control regions in data to estimate reducible backgrounds and to validate simulations used to estimate irreducible background processes. To maximize sensitivity to a broad range of possible signal models, 46 exclusive signal regions are defined. No significant deviation from the expected standard model background is observed in any of these signal regions.

The results are interpreted using a simplified gluino-pair production model that features cascade decays producing four top quarks and two neutralinos. In this model, gluinos with a mass up to 1610 GeV are excluded in the case of a massless LSP. The maximum excluded LSP mass is 900 GeV. This represents an improvement of approximately 435 and 250 GeV, respectively, compared to the exclusion limit set in a similar search based on data collected with the CMS detector in 2015, corresponding to an integrated luminosity of $2.3\,\text{fb}^{-1}$ [39].

For the simplified model of gluino-gluino production with decay to light-flavor quark jets, two vector bosons and neutralinos, gluino masses up to 1160 GeV and neutralino masses up to 680 GeV can be excluded. The limit on gluino and neutralino masses extends the corresponding limit from the previous analysis by about 335 and 180 GeV, respectively.

For a simplified model of bottom squark pair production decaying to top quarks, W bosons and neutralinos, bottom squark masses up to 840 GeV are excluded for a low mass chargino, while chargino masses are excluded up to 750 GeV. These extend the previous limits by 380 GeV for each particle.

Finally, for a simplified heavy top squark pair production model with further decays to two top quarks, Higgs or Z bosons, and neutralinos, the $\widetilde{t}_2$ mass is excluded up to 720, 780, and 710 GeV for models with an exclusive $\widetilde{t}_2 \to \widetilde{t}_1 H$ decay, an exclusive $\widetilde{t}_2 \to \widetilde{t}_1 Z$ decay, and an equally probable mix of those two decays, while the $\widetilde{t}_1$ mass is excluded up to 430, 540, and 450 GeV for the same branching fractions. This significantly improves the results obtained with the 8 TeV dataset [36].

Acknowledgments

We congratulate our colleagues in the CERN accelerator departments for the excellent performance of the LHC and thank the technical and administrative staffs at CERN and at other CMS institutes for their contributions to the success of the CMS effort. In addition, we gratefully acknowledge the computing centres and personnel of the Worldwide LHC Computing Grid for delivering so effectively the computing infrastructure essential to our analyses. Finally, we acknowledge the enduring support for the construction and operation of the LHC and the CMS detector provided by the following funding agencies: BMWFW and FWF (Austria); FNRS and FWO (Belgium); CNPq, CAPES, FAPERJ, and FAPESP (Brazil); MES (Bulgaria); CERN; CAS, MoST, and NSFC (China); COLCIENCIAS (Colombia); MSES and CSF (Croatia); RPF (Cyprus); SENESCYT (Ecuador); MoER, ERC IUT, and ERDF (Estonia); Academy of Finland, MEC, and HIP (Finland); CEA and CNRS/IN2P3 (France); BMBF, DFG, and HGF (Germany); GSRT (Greece); OTKA and NIH (Hungary); DAE and DST (India); IPM (Iran); SFI (Ireland); INFN (Italy); MSIP and NRF (Republic of Korea); LAS (Lithuania); MOE and UM (Malaysia); BUAP, CINVESTAV, CONACYT, LNS, SEP, and UASLP-FAI (Mexico); MBIE (New

Zealand); PAEC (Pakistan); MSHE and NSC (Poland); FCT (Portugal); JINR (Dubna); MON, RosAtom, RAS, RFBR and RAEP (Russia); MESTD (Serbia); SEIDI, CPAN, PCTI and FEDER (Spain); Swiss Funding Agencies (Switzerland); MST (Taipei); ThEPCenter, IPST, STAR, and NSTDA (Thailand); TUBITAK and TAEK (Turkey); NASU and SFFR (Ukraine); STFC (United Kingdom); DOE and NSF (USA).

Individuals have received support from the Marie-Curie programme and the European Research Council and Horizon 2020 Grant, contract No. 675440 (European Union); the Leventis Foundation; the A. P. Sloan Foundation; the Alexander von Humboldt Foundation; the Belgian Federal Science Policy Office; the Fonds pour la Formation à la Recherche dans l'Industrie et dans l'Agriculture (FRIA-Belgium); the Agentschap voor Innovatie door Wetenschap en Technologie (IWT-Belgium); the Ministry of Education, Youth and Sports (MEYS) of the Czech Republic; the Council of Science and Industrial Research, India; the HOMING PLUS programme of the Foundation for Polish Science, cofinanced from European Union, Regional Development Fund, the Mobility Plus programme of the Ministry of Science and Higher Education, the National Science Center (Poland), contracts Harmonia 2014/14/M/ST2/00428, Opus 2014/13/B/ST2/02543, 2014/15/B/ST2/03998, and 2015/19/B/ST2/02861, Sonata-bis 2012/07/E/ST2/01406; the National Priorities Research Program by Qatar National Research Fund; the Programa Severo Ochoa del Principado de Asturias; the Thalis and Aristeia programmes cofinanced by EU-ESF and the Greek NSRF; the Rachadapisek Sompot Fund for Postdoctoral Fellowship, Chulalongkorn University and the Chulalongkorn Academic into Its 2nd Century Project Advancement Project (Thailand); the Welch Foundation, contract C-1845; and the Weston Havens Foundation (USA).

References

[1] O. J. P. Eboli, R. Zukanovich Funchal, and T. L. Lungov, "Multilepton signatures for leptoquarks", *Phys. Rev. D* **59** (1999) 035002, doi:10.1103/PhysRevD.59.035002, arXiv:hep-ph/9808288.

[2] N. Craig et al., "Searching for $t \to ch$ with multi-leptons", *Phys. Rev. D* **86** (2012) 075002, doi:10.1103/PhysRevD.86.075002, arXiv:1207.6794.

[3] N. Craig et al., "Heavy Higgs bosons at low tan β: from the LHC to 100 TeV", *JHEP* **01** (2017) 018, doi:10.1007/JHEP01(2017)018, arXiv:1605.08744.

[4] N. Craig et al., "Multi-lepton signals of multiple Higgs bosons", *JHEP* **02** (2013) 033, doi:10.1007/JHEP02(2013)033, arXiv:1210.0559.

[5] C.-S. Chen and Y.-J. Zheng, "LHC signatures for the cascade seesaw mechanism", *PTEP* **2015** (2015) 103B02, doi:10.1093/ptep/ptv134, arXiv:1312.7207.

[6] A. Lazopoulos, T. McElmurry, K. Melnikov, and F. Petriello, "Next-to-leading order QCD corrections to $t\bar{t}Z$ production at the LHC", *Phys. Lett. B* **666** (2008) 62, doi:10.1016/j.physletb.2008.06.073, arXiv:0804.2220.

[7] A. Kardos, Z. Trocsanyi, and C. Papadopoulos, "Top quark pair production in association with a Z-boson at NLO accuracy", *Phys. Rev. D* **85** (2012) 054015, doi:10.1103/PhysRevD.85.054015, arXiv:1111.0610.

[8] J. M. Campbell and R. K. Ellis, "$t\bar{t}W^{\pm}$ production and decay at NLO", *JHEP* **07** (2012) 052, doi:10.1007/JHEP07(2012)052, arXiv:1204.5678.

[9] J. Campbell, R. K. Ellis, and R. Röntsch, "Single top production in association with a Z boson at the LHC", *Phys. Rev. D* **87** (2013) 114006, doi:10.1103/PhysRevD.87.114006, arXiv:1302.3856.

[10] A. Kulesza, L. Motyka, T. Stebel, and V. Theeuwes, "Soft gluon resummation for associated $t\bar{t}H$ production at the LHC", *JHEP* **03** (2016) 065, doi:10.1007/JHEP03(2016)065, arXiv:1509.02780.

[11] A. Broggio et al., "Associated production of a top pair and a Higgs boson beyond NLO", *JHEP* **03** (2016) 124, doi:10.1007/JHEP03(2016)124, arXiv:1510.01914.

[12] M. Grazzini, S. Kallweit, D. Rathlev, and M. Wiesemann, "$W^{\pm}Z$ production at the LHC: fiducial cross sections and distributions in NNLO QCD", *JHEP* **05** (2017) 139, doi:10.1007/JHEP05(2017)139, arXiv:1703.09065.

[13] F. Cascioli et al., "ZZ production at hadron colliders in NNLO QCD", *Phys. Lett. B* **735** (2014) 311, doi:10.1016/j.physletb.2014.06.056, arXiv:1405.2219.

[14] F. Caola, K. Melnikov, R. Röntsch, and L. Tancredi, "QCD corrections to ZZ production in gluon fusion at the LHC", *Phys. Rev. D* **92** (2015) 094028, doi:10.1103/PhysRevD.92.094028, arXiv:1509.06734.

[15] J. M. Campbell, R. K. Ellis, M. Czakon, and S. Kirchner, "Two loop correction to interference in $gg \to ZZ$", *JHEP* **08** (2016) 011, doi:10.1007/JHEP08(2016)011, arXiv:1605.01380.

[16] T. Binoth, G. Ossola, C. G. Papadopoulos, and R. Pittau, "NLO QCD corrections to tri-boson production", *JHEP* **06** (2008) 082, doi:10.1088/1126-6708/2008/06/082, arXiv:0804.0350.

[17] D. T. Nhung, L. D. Ninh, and M. M. Weber, "NLO corrections to WWZ production at the LHC", *JHEP* **12** (2013) 096, doi:10.1007/JHEP12(2013)096, arXiv:1307.7403.

[18] S. Yong-Bai et al., "NLO QCD + NLO EW corrections to WZZ productions with leptonic decays at the LHC", *JHEP* **10** (2015) 186, doi:10.1007/JHEP10(2015)186, arXiv:1507.03693. [Erratum: *JHEP* **10** (2016) 156, doi:10.1007/JHEP10(2016)156].

[19] Y.-B. Shen et al., "NLO QCD and electroweak corrections to WWW production at the LHC", *Phys. Rev. D* **95** (2017) 073005, doi:10.1103/PhysRevD.95.073005, arXiv:1605.00554.

[20] W. Hong et al., "NLO QCD + EW corrections to ZZZ production with subsequent leptonic decays at the LHC", *J. Phys. G* **43** (2016) 115001, doi:10.1088/0954-3899/43/11/115001, arXiv:1610.05876.

[21] ATLAS Collaboration, "Measurement of the $t\bar{t}Z$ and $t\bar{t}W$ production cross sections in multilepton final states using 3.2 fb^{-1} of pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector", *Eur. Phys. J. C* **77** (2017) 40, doi:10.1140/epjc/s10052-016-4574-y, arXiv:1609.01599.

[22] ATLAS Collaboration, "Measurement of the $W^{\pm}Z$ boson pair-production cross section in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector", *Phys. Lett. B* **762** (2016) 1, doi:10.1016/j.physletb.2016.08.052, arXiv:1606.04017.

[23] CMS Collaboration, "Measurement of the WZ production cross section in pp collisions at $\sqrt{s} = 13$ TeV", *Phys. Lett. B* **766** (2017) 268, doi:10.1016/j.physletb.2017.01.011, arXiv:1607.06943.

[24] ATLAS Collaboration, "Measurement of the ZZ production cross section in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector", *Phys. Rev. Lett.* **116** (2016) 101801, doi:10.1103/PhysRevLett.116.101801, arXiv:1512.05314.

[25] CMS Collaboration, "Measurement of the ZZ production cross section and $Z \to \ell^+\ell^-\ell'^+\ell'^-$ branching fraction in pp collisions at $\sqrt{s} = 13$ TeV", *Phys. Lett. B* **763** (2016) 280, doi:10.1016/j.physletb.2016.10.054, arXiv:1607.08834.

[26] P. Ramond, "Dual theory for free fermions", *Phys. Rev. D* **3** (1971) 2415, doi:10.1103/PhysRevD.3.2415.

[27] Yu. A. Golfand and E. P. Likhtman, "Extension of the algebra of Poincare group generators and violation of P invariance", *JETP Lett.* **13** (1971) 323. [*Pisma Zh. Eksp. Teor. Fiz.* **13** (1971) 452].

[28] A. Neveu and J. H. Schwarz, "Factorizable dual model of pions", *Nucl. Phys. B* **31** (1971) 86, doi:10.1016/0550-3213(71)90448-2.

[29] D. V. Volkov and V. P. Akulov, "Possible universal neutrino interaction", *JETP Lett.* **16** (1972) 438. [*Pisma Zh. Eksp. Teor. Fiz.* **16** (1972) 621].

[30] J. Wess and B. Zumino, "A lagrangian model invariant under supergauge transformations", *Phys. Lett. B* **49** (1974) 52, doi:10.1016/0370-2693(74)90578-4.

[31] J. Wess and B. Zumino, "Supergauge transformations in four-dimensions", *Nucl. Phys. B* **70** (1974) 39, doi:10.1016/0550-3213(74)90355-1.

[32] P. Fayet, "Supergauge invariant extension of the Higgs mechanism and a model for the electron and its neutrino", *Nucl. Phys. B* **90** (1975) 104, doi:10.1016/0550-3213(75)90636-7.

[33] H. P. Nilles, "Supersymmetry, supergravity and particle physics", *Phys. Rept.* **110** (1984) 1, doi:10.1016/0370-1573(84)90008-5.

[34] S. P. Martin, "A supersymmetry primer", *Adv. Ser. Direct. High Energy Phys.* **21** (2010) 1, doi:10.1142/9789814307505_0001, arXiv:hep-ph/9709356.

[35] D. Alves et al., "Simplified models for LHC new physics searches", *J. Phys. G* **39** (2012) 105005, doi:10.1088/0954-3899/39/10/105005, arXiv:1105.2838.

[36] CMS Collaboration, "Search for top-squark pairs decaying into Higgs or Z bosons in pp collisions at $\sqrt{s} = 8$ TeV", *Phys. Lett. B* **736** (2014) 371, doi:10.1016/j.physletb.2014.07.053, arXiv:1405.3886.

[37] ATLAS Collaboration, "Search for direct top squark pair production in events with a Higgs or Z boson, and missing transverse momentum in $\sqrt{s} = 13$ TeV pp collisions with the ATLAS detector", *JHEP* **08** (2017) 006, doi:10.1007/JHEP08(2017)006, arXiv:1706.03986.

[38] ATLAS Collaboration, "Search for supersymmetry in final states with two same-sign or three leptons and jets using 36 fb^{-1} of $\sqrt{s} = 13$ TeV pp collision data with the ATLAS detector", *JHEP* **09** (2017) 084, doi:10.1007/JHEP09(2017)084, arXiv:1706.03731.

[39] CMS Collaboration, "Search for supersymmetry with multiple charged leptons in proton-proton collisions at $\sqrt{s} = 13$ TeV", *Eur. Phys. J. C* **77** (2017) 635, doi:10.1140/epjc/s10052-017-5182-1, arXiv:1701.06940.

[40] CMS Collaboration, "Performance of photon reconstruction and identification with the CMS detector in proton-proton collisions at $\sqrt{s} = 8$ TeV", *JINST* **10** (2015) P08010, doi:10.1088/1748-0221/10/08/P08010, arXiv:1502.02702.

[41] CMS Collaboration, "Performance of CMS muon reconstruction in pp collision events at $\sqrt{s} = 7$ TeV", *JINST* **7** (2012) P10002, doi:10.1088/1748-0221/7/10/P10002, arXiv:1206.4071.

[42] CMS Collaboration, "The CMS trigger system", *JINST* **12** (2017) P01020, doi:10.1088/1748-0221/12/01/P01020, arXiv:1609.02366.

[43] CMS Collaboration, "The CMS experiment at the CERN LHC", *JINST* **3** (2008) S08004, doi:10.1088/1748-0221/3/08/S08004.

[44] CMS Collaboration, "Particle-flow reconstruction and global event description with the CMS detector", *JINST* **12** (2017) P10003, doi:10.1088/1748-0221/12/10/P10003, arXiv:1706.04965.

[45] M. Cacciari, G. P. Salam, and G. Soyez, "The anti-k_t jet clustering algorithm", *JHEP* **04** (2008) 063, doi:10.1088/1126-6708/2008/04/063, arXiv:0802.1189.

[46] M. Cacciari, G. P. Salam, and G. Soyez, "FastJet user manual", *Eur. Phys. J. C* **72** (2012) 1896, doi:10.1140/epjc/s10052-012-1896-2, arXiv:1111.6097.

[47] CMS Collaboration, "Determination of jet energy calibration and transverse momentum resolution in CMS", *JINST* **6** (2011) P11002, doi:10.1088/1748-0221/6/11/P11002, arXiv:1107.4277.

[48] CMS Collaboration, "Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV", *JINST* **12** (2016) P02014, doi:10.1088/1748-0221/12/02/P02014, arXiv:1607.03663.

[49] CMS Collaboration, "Identification of b-quark jets with the CMS experiment", *JINST* **8** (2013) P04013, doi:10.1088/1748-0221/8/04/P04013, arXiv:1211.4462.

[50] CMS Collaboration, "Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV", (2017). arXiv:1712.07158. Submitted to *JINST*.

[51] CMS Collaboration, "Performance of the CMS missing transverse momentum reconstruction in pp data at $\sqrt{s} = 8$ TeV", *JINST* **10** (2015) P02006, doi:10.1088/1748-0221/10/02/P02006, arXiv:1411.0511.

[52] CMS Collaboration, "Performance of missing energy reconstruction in 13 TeV pp collision data using the CMS detector", CMS Physics Analysis Summary CMS-PAS-JME-16-004, 2016.

[53] CMS Collaboration, "Performance of electron reconstruction and selection with the CMS detector in proton-proton collisions at $\sqrt{s} = 8$ TeV", *JINST* **10** (2015) P06005, doi:10.1088/1748-0221/10/06/P06005, arXiv:1502.02701.

[54] CMS Collaboration, "Search for new physics in same-sign dilepton events in proton-proton collisions at $\sqrt{s} = 13$ TeV", *Eur. Phys. J. C* **76** (2016) 439, doi:10.1140/epjc/s10052-016-4261-z, arXiv:1605.03171.

[55] J. Alwall et al., "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations", *JHEP* **07** (2014) 079, doi:10.1007/JHEP07(2014)079, arXiv:1405.0301.

[56] S. Alioli, P. Nason, C. Oleari, and E. Re, "A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX", *JHEP* **06** (2010) 043, doi:10.1007/JHEP06(2010)043, arXiv:1002.2581.

[57] H. B. Hartanto, B. Jager, L. Reina, and D. Wackeroth, "Higgs boson production in association with top quarks in the POWHEG BOX", *Phys. Rev. D* **91** (2015) 094003, doi:10.1103/PhysRevD.91.094003, arXiv:1501.04498.

[58] T. Melia, P. Nason, R. Rontsch, and G. Zanderighi, "W^+W^-, WZ and ZZ production in the POWHEG BOX", *JHEP* **11** (2011) 078, doi:10.1007/JHEP11(2011)078, arXiv:1107.5051.

[59] P. Nason and G. Zanderighi, "W^+W^-, WZ and ZZ production in the POWHEG-BOX-V2", *Eur. Phys. J. C* **74** (2014) 2702, doi:10.1140/epjc/s10052-013-2702-5, arXiv:1311.1365.

[60] NNPDF Collaboration, "Parton distributions for the LHC Run II", *JHEP* **04** (2015) 040, doi:10.1007/JHEP04(2015)040, arXiv:1410.8849.

[61] T. Sjöstrand, S. Mrenna, and P. Z. Skands, "A brief introduction to PYTHIA 8.1", *Comput. Phys. Commun.* **178** (2008) 852, doi:10.1016/j.cpc.2008.01.036, arXiv:0710.3820.

[62] P. Skands, S. Carrazza, and J. Rojo, "Tuning PYTHIA 8.1: the Monash 2013 tune", *Eur. Phys. J. C* **74** (2014) 3024, doi:10.1140/epjc/s10052-014-3024-y, arXiv:1404.5630.

[63] CMS Collaboration, "Event generator tunes obtained from underlying event and multiparton scattering measurements", *Eur. Phys. J. C* **76** (2016) 155, doi:10.1140/epjc/s10052-016-3988-x, arXiv:1512.00815.

[64] J. Alwall et al., "Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions", *Eur. Phys. J. C* **53** (2008) 473, doi:10.1140/epjc/s10052-007-0490-5, arXiv:0706.2569.

[65] R. Frederix and S. Frixione, "Merging meets matching in MC@NLO", *JHEP* **12** (2012) 061, doi:10.1007/JHEP12(2012)061, arXiv:1209.6215.

[66] GEANT4 Collaboration, "GEANT4—a simulation toolkit", *Nucl. Instrum. Meth. A* **506** (2003) 250, doi:10.1016/S0168-9002(03)01368-8.

[67] S. Abdullin et al., "The fast simulation of the CMS detector at LHC", *J. Phys. Conf. Ser.* **331** (2011) 032049, doi:10.1088/1742-6596/331/3/032049.

[68] W. Beenakker, R. Hopker, M. Spira, and P. M. Zerwas, "Squark and gluino production at hadron colliders", *Nucl. Phys. B* **492** (1997) 51, doi:10.1016/S0550-3213(97)80027-2, arXiv:hep-ph/9610490.

[69] A. Kulesza and L. Motyka, "Threshold resummation for squark-antisquark and gluino-pair production at the LHC", *Phys. Rev. Lett.* **102** (2009) 111802, doi:10.1103/PhysRevLett.102.111802, arXiv:0807.2405.

[70] A. Kulesza and L. Motyka, "Soft gluon resummation for the production of gluino-gluino and squark-antisquark pairs at the LHC", *Phys. Rev. D* **80** (2009) 095004, doi:10.1103/PhysRevD.80.095004, arXiv:0905.4749.

[71] W. Beenakker et al., "Soft-gluon resummation for squark and gluino hadroproduction", *JHEP* **12** (2009) 041, doi:10.1088/1126-6708/2009/12/041, arXiv:0909.4418.

[72] W. Beenakker et al., "Squark and gluino hadroproduction", *Int. J. Mod. Phys. A* **26** (2011) 2637, doi:10.1142/S0217751X11053560, arXiv:1105.1110.

[73] C. Borschensky et al., "Squark and gluino production cross sections in pp collisions at $\sqrt{s}$ = 13, 14, 33 and 100 TeV", *Eur. Phys. J. C* **74** (2014) 3174, doi:10.1140/epjc/s10052-014-3174-y, arXiv:1407.5066.

[74] Particle Data Group, "Review of particle physics", *Chin. Phys. C* **40** (2016) 100001, doi:10.1088/1674-1137/40/10/100001.

[75] T. Junk, "Confidence level computation for combining searches with small statistics", *Nucl. Instrum. Meth. A* **434** (1999) 435, doi:10.1016/S0168-9002(99)00498-2, arXiv:hep-ex/9902006.

[76] A. L. Read, "Presentation of search results: the CL_s technique", *J. Phys. G* **28** (2002) 2693, doi:10.1088/0954-3899/28/10/313.

[77] ATLAS and CMS Collaborations, "Procedure for the LHC higgs boson search combination in summer 2011", Technical Report ATL-PHYS-PUB-2011-11, CMS-NOTE-2011-005, CERN, 2011.

[78] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, "Asymptotic formulae for likelihood-based tests of new physics", *Eur. Phys. J. C* **71** (2011) 1554, doi:10.1140/epjc/s10052-011-1554-0, arXiv:1007.1727. [Erratum: *Eur. Phys. J. C* **73** (2013) 2501, doi:10.1140/epjc/s10052-013-2501-z].

6.1 SUSY SIGNAL MODELS

Simplified SUSY models are used as benchmark processes in order to optimize the event selection and to interpret the results in terms of exclusion limits on the masses of the SUSY particles in a given model. Of particular interests are gluino pair and 3rd generation squark pair production with their subsequent cascade decays. All considered models assume R parity conservation, which means each model will have an LSP in the final state. The LSP escapes detection and gives rise to missing transverse momentum in the event. In this section, we will discuss the features of the simplified models considered in the analysis.

6.1.1 *T1tttt: A model with four top quarks in the final state*

The first considered model is gluino pair production with the subsequent decay of each gluino via an intermediate virtual top squark into a top quark pair and a neutralino. In most of the phase space, this final state would result in a high jet and b jet multiplicity, as well as high H_{T} and p_{T}^{miss}. The multilepton signature constitutes only about 7% of the 4 top-quark decays but it greatly profits from the reduced SM backgrounds and therefore provides a competitive sensitivity compared to other signatures. The model is denoted as T1tttt and is depicted in Fig. 6.1.

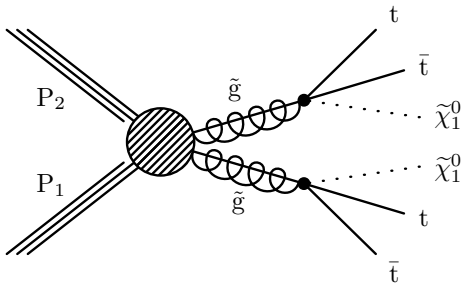


Figure 6.1: Gluino pair production decaying into four top quarks and two neutralinos via an intermediate virtual $\widetilde{t}$ sparticle (T1tttt).

6.1.2 *T6ttWW: A model with four W bosons and two b quarks in the final state*

One of the models considered to represent 3$^{\mathrm{rd}}$ generation squark production is a $\widetilde{b}$-pair production process where each $\widetilde{b}$ undergoes a cascade decay through an on-shell chargino (see Fig. 6.2). In this model, the mass of the LSP is set to 50 GeV. This process has only 2 b quarks in the final state. In this search the masses of b squark are about 750-850 GeV which is significantly lower than the gluino masses this analysis sensitive to. This leads to the fact that in T6ttWW model the final state contains rather soft jets as opposed to those in T1tttt. Therefore the sensitivity of the fully hadronic searches for T6ttWW is reduced and the sensitivity reached by the multilepton signature becomes more important.

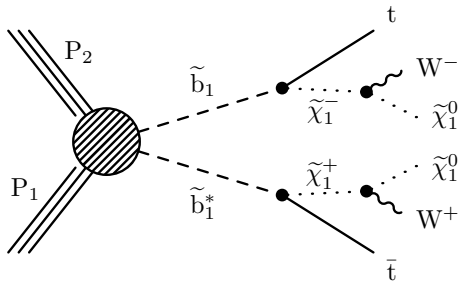


Figure 6.2: Bottom squark pair production decaying into two top quarks, two W bosons and two neutralinos (T6ttWW).

6.1.3 *T6ttHZ: A model with BEH and Z bosons*

Another model considered in this search consists of a $\tilde{t}_2$-pair production process where each $\tilde{t}_2$ squark undergoes a cascade decay through a $\tilde{t}_1$ (see Fig. 6.3). This process has at least 2 b quarks in the final state, along with two prompt leptons from the top quarks. Additional leptons may be provided by the decays of the BEH and Z bosons, or also b jets in the case of the BEH decay. In this model, the mass difference between the $\tilde{t}_1$ and the LSP is set to 175 GeV, the $\tilde{t}_1$ decays via a top quark to LSP, and the $\tilde{t}_2$ decays via a Z or BEH boson to $\tilde{t}_1$. We consider the reference values $\mathcal{B}(\tilde{t}_2 \to \tilde{t}_1 Z) = 0$, 50, and 100%; the sensitivity is diminished for the $\tilde{t}_1 H$ final state because of the additional branching factors for BEH cascade decays to electrons or muons via gauge bosons or τ leptons.

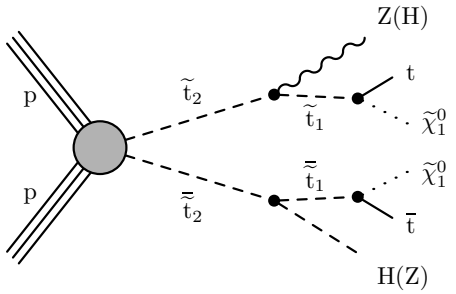


Figure 6.3: Top squark pair production decaying into BEH/Z bosons, two top quarks and two neutralinos (T6ttHZ).

6.1.4 *T5qqqqVV: A model without b quarks in the final state*

The Fig. 6.4 shows another simplified model with gluino pair production, referred to as T5qqqqVV. Here, the gluinos decay into light flavor quarks and either the lightest chargino ($\tilde{\chi}_1^\pm$) or the second lightest neutralino ($\tilde{\chi}_2^0$). The probability for the decay to proceed via the $\tilde{\chi}_1^-$, $\tilde{\chi}_1^+$, or $\tilde{\chi}_2^0$ is taken to be 1/3 for each case. Subsequently, the chargino (neutralino) decays into a W(Z) boson and the LSP. In this scenario, the second neutralino $\tilde{\chi}_2^0$ and chargino are assumed to be mass-degenerate, with masses equal to $0.5(m_{\tilde{g}} + m_{\tilde{\chi}_1^0})$.

Unlike previously described models, this model does not produce b jets in the final state. Instead, a pair of opposite-charge, same-flavor (OCSF) leptons is produced if the Z boson decays leptonically. Depending on the mass splitting between the $\widetilde{\chi}_2^0/\widetilde{\chi}_1^\pm$ and $\widetilde{\chi}_1^0$ the leptons in the final state can be either soft or hard. Different combinations of vector bosons give rise to three distinct signal topologies. Assuming equal probabilities for the three decay channels relative production probabilities are 4/9 and 1/9 for the topology featuring WZ and ZZ production, respectively. While these two topologies can produce up to three and four leptons respectively, the WW topology, produced with a probability of 4/9 yields at most two leptons if both W bosons decay leptonically.

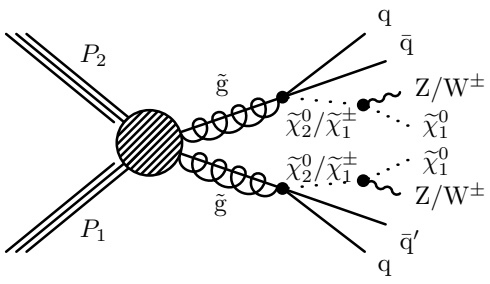


Figure 6.4: Gluino or squark production decaying into leptons and light quarks (T5qqqqVV).

6.2 TRIGGER AND OBJECT SELECTION

In this section, we briefly discuss the trigger, lepton and jet selections used in the analysis.

6.2.1 *Triggers*

The analysis targets events with at least three well-identified muons or electrons in any combination. These events are required to pass trigger selection criteria that target dilepton and multilepton events. The first set of triggers, referred to as isolated dilepton triggers, selects events with at least two loosely isolated leptons. On top of isolation, the longitudinal distance between two leptons is added to select leptons from the same vertex. The isolated dilepton triggers employed in this analysis impose p_T thresholds of 17 GeV and 8 GeV on the highest-p_T (leading) and the second highest-p_T (sub-leading) muon. The corresponding thresholds for electrons are higher — 23 GeV and 12 GeV for the leading and the sub-leading electron, respectively. The second set of dilepton triggers places no requirements on the isolation. Instead, these triggers require a minimum amount of hadronic activity reconstructed in the trigger to be greater than 300 GeV. The lepton threshold for this type of triggers is set to 8 GeV both for electrons and muons. All trigger names are summarized in Table B.1.

For the dilepton triggers considered here, the trigger efficiency can be factorized as a product of the individual efficiencies for the trigger to correctly recognize the leading lepton, the sub-leading lepton and, in case of the non-isolated triggers, the hadronic activity separately. These efficiencies are measured in data using the tag-and-probe technique discussed earlier. Events entering the denominator of the efficiency are selected by a single lepton trigger that imposes more stringent requirements on the so-called tag lepton than the dilepton trigger requirement under study. Additionally, the tag lepton

has to pass the specific lepton selection that will be described in subsection 6.2.2. A second lepton in the event, which also has to pass the lepton selection with the p_{T} requirement omitted, serves as the probe lepton. This allows measuring the factorized efficiency for triggering the event as a function of p_{T} and η of the probe lepton. An example of measured efficiencies is shown in Fig. 6.5.

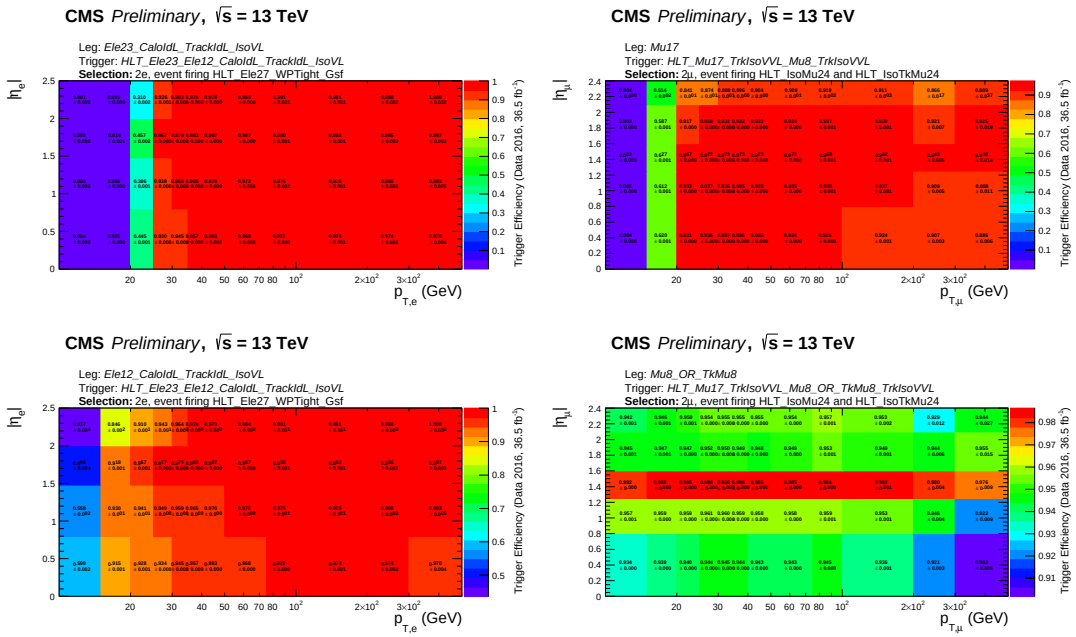


Figure 6.5: Efficiency (η,p_{T}) map for the leading (top) and sub-leading (bottom) legs of the dielectron (left) and dimuon (right) triggers in the full 2016 data set. The figures are made by Laurent Thomas.

The efficiencies have been measured to be 90% or better for leading and sub-leading leptons of both flavors and all trigger paths considered in this analysis. The redundancy of triggering events with at least three leptons with dilepton triggers thus entails trigger efficiencies above 97% if the leptons have sufficiently large transverse momenta. The lower lepton p_{T} threshold of the non-isolated triggers allows decreasing the offline p_{T} requirement for the offline lepton selection.

In addition to the signal triggers, a set of single lepton triggers that impose no isolation requirement on the lepton are used in the analysis. Events selected by these auxiliary triggers are used to measure the lepton misidentification rate that is used for the nonprompt lepton background estimation, as will be discussed in the section 6.4. Owing to the large number of events that can pass these triggers, their rates have been decreased by applying different pre-scale factors, resulting in lower effective integrated luminosities, compared to the signal triggers.

6.2.2 *Lepton selection*

Robust identification of muons and electrons is important to achieve a high selection efficiency in events with multiple leptons. On the other hand, prompt muons and electrons need to be discriminated against their nonprompt counterparts. Two different sets of quality requirements are used to form selection criteria for each of the lepton flavors, differing in the purity of contained prompt leptons. A so-called tight selection

with stringent quality requirements aims at reducing the contribution from nonprompt leptons as much as possible while maintaining a high selection efficiency for prompt leptons. Leptons satisfying these selection criteria form the basis for identifying events with multiple leptons for a SUSY particles search. A second set of selection criteria with relaxed quality requirements is referred to as loose selection. Events with leptons passing this selection are used for an estimation of the nonprompt lepton background. Even more, relaxed requirements in identification and impact parameters are used for muons to avoid double counting of electrons and muons. The selections are inclusive, i.e., the more stringent selection contains a subset of leptons that pass the respective looser selection.

Two different working points of the MVA are used for electrons in the tight and the loose selection. The more rigorous working point is tuned to select prompt electrons with an efficiency of about 90% and has an acceptance of nonprompt electrons of 10% (20%) in the barrel (endcap) region. The working point used for the loose electron selection has been optimized with regard to the performance of the nonprompt background estimation, which will be discussed in Section 6.4. A summary of all lepton identification criteria for the lepton selections used in the analysis is given in Table 6.1 for muons and electrons. The difference between the loose and the tight selection is the more stringent requirement on isolation and stricter demands on the electron identification in terms of electron MVA value.

Electrons		
Criterion	Loose	Tight
$p_T > 10$ GeV	✓	✓
$\|\eta\| < 2.5$	✓	✓
$\|d_{xy}\| < 0.05$ cm	✓	✓
$\|d_z\| < 0.1$ cm	✓	✓
$SIP_{3D} < 4$	✓	✓
conversion rejection	✓	✓
no missing pixel hits	✓	✓
trigger emulation	✓	✓
electron MVA loose	✓	✓
electron MVA tight	-	✓
multi-isolation	-	✓

Muons		
Criterion	Loose	Tight
$p_T > 10$ GeV	✓	✓
$\|\eta\| < 2.4$	✓	✓
$\|d_{xy}\| < 0.05$ cm	✓	✓
$\|d_z\| < 0.1$ cm	✓	✓
$SIP_{3D} < 4$	✓	✓
Loose ID	✓	✓
Medium ID	✓	✓
multi-isolation	-	✓

Table 6.1: List of all quality criteria for electrons and muons for passing the loose and the tight selection in the SUSY analysis.

6.2.3 *Jet selection*

Jets used in this analysis are asked to pass "Loose" identification criteria, have to satisfy $p_T > 30$ GeV requirement, and must reside within $|\eta| < 2.4$. Double counting with leptons is avoided by discarding the jet closest to a muon or electron passing the loose selection if the separation between the jet and the lepton is smaller than 0.4 in ΔR. Same requirements are applied for b jets, apart from a relaxed to 25 GeV requirement

for the p_{T}. The latter enhances the discrimination between topologies with or without b jets in the final state. The CSVv2 b tagging algorithm is used to identify b jets in this analysis.

6.3 EVENT SELECTION AND SEARCH STRATEGY

Any SUSY search faces the challenge that the mass hierarchy of the predicted SUSY particles and their experimental signature are inherently unknown. The mass hierarchy of the SUSY particles directly govern the kinematic properties of the final state. For example, in T5qqqqVV model the mass difference between gluino and chargino or $\widetilde{\chi}_2^0$ determines the amount of H_{T}, while the difference between the masses of gluino and LSP determines the size of p_{T}^{miss}. Furthermore, even if these mass hierarchies are fixed the final state particles can be very different in terms of number of b-tagged jets, presence of on-shell W and Z bosons, etc. In order to maximize the sensitivity of the search, the presented analysis follows an inclusive approach to achieve sensitivity for a wide range of BSM physics processes that produce three or more prompt leptons.

Up to this moment, a selection on trigger level is applied to filter out events of interest out of the huge amount of proton-proton collisions recorded by the CMS experiment. At first, we discuss the baseline selection, which aims to reduce the rate of background processes as much as possible while maintaining a high signal efficiency for all the simplified SUSY models described in Section 6.1. In order to further enhance the signal-to-background ratio for different models, selected events are categorized into different signal regions (SRs). In these SRs, SUSY physics processes would manifest themselves as notable excess of data over the SM background prediction.

6.3.1 *Baseline selection*

The selection requires the presence of at least three well-identified leptons in the event. The leptons must satisfy p_{T} thresholds that depend on the lepton flavor and the amount of hadronic activity in the event. For events with low hadronic activity ($H_{\mathrm{T}} < 400$ GeV), the leading electron(muon) must satisfy $p_{\mathrm{T}} > 25(20)$ GeV and sub-leading electrons (muons) must satisfy $p_{\mathrm{T}} > 15(10)$ GeV. In events with high hadronic activity ($H_{\mathrm{T}} > 400$ GeV), the thresholds are relaxed to 15(10) GeV for the leading electrons (muons). The lowest-p_{T} (trailing) lepton must have $p_{\mathrm{T}} > 10$ GeV in all cases. Invariant mass of OCSF lepton pairs are required to be greater than 12 GeV to suppress DY and quarkonium processes.

Additional rejection of DY events with higher invariant mass is achieved by selecting only events with $p_{\mathrm{T}}^{miss} > 50$ GeV, as such events do not involve genuine p_{T}^{miss}. In the considered simplified models we expect a significant number of jets in the final state, therefore we require to have at least 2 jets in any selected event ($N_{\mathrm{j}} \geq 2$). Given the definition of H_{T}, this requirement is equivalent to $H_{\mathrm{T}} > 60$ GeV. Selecting events with at least 2 jets also rejects a large portion of the diboson processes, since no jets are produced in this process at LO. Additional rejection of DY events is achieved by increasing the minimum requirement for p_{T}^{miss} from 50 GeV to 70 GeV for events with a Z boson, small b jet multiplicities, low H_{T}, and low p_{T}^{miss}. All these selection criteria are summarized in Table 6.2.

To maximize the sensitivity for signal processes with or without on-shell Z bosons, events are split into two exclusive regions according to invariant mass of OCSF pair, $\mathrm{m}_{\ell\ell}$:

Number of selected leptons	$\geq$3
N_{j}	$\geq$2
p_{T}^{miss}, GeV	>50 (70 in low N_{b}, p_{T}^{miss} and H_{T} category)
$\mathrm{m}_{\ell\ell}$, GeV	>12

Table 6.2: Summary of all requirements used in the baseline selection criteria.

events with $\mathrm{m}_{\ell\ell}$ 15 GeV close to a mass of a Z boson are denoted as on-Z events. All others, including those without any OCSF dilepton pair, are called off-Z events.

The background rejection achieved by each baseline selection criterion in on-Z and off-Z categories is presented in Tables 6.4 and 6.3 for different background processes. The requirements for N_{j} and p_{T}^{miss} efficiently suppress the background from the DY process. The dominant background sources in the on-Z and off-Z region are $\mathrm{t\bar{t}}$, WZ, $\mathrm{t\bar{t}Z}$, and $\mathrm{t\bar{t}W}$ processes.

Process	DY	$\mathrm{t\bar{t}}$	WZ	$\mathrm{t\bar{t}Z}$	$\mathrm{t\bar{t}W}$
$\geq$ 3 leptons, $p_{\mathrm{T}} > 10$ GeV	3120.5 ± 142.2	1107.3 ± 24.2	2912.4 ± 15.8	396.8 ± 0.8	82.7 ± 0.3
$m_{\ell\ell} \geq 12$ GeV	3120.5 ± 142.2	1107.3 ± 24.2	2912.4 ± 15.8	396.8 ± 0.8	82.7 ± 0.3
lepton p_{T} selection	3087.4 ± 141.6	1092.6 ± 24.2	2893.4 ± 15.8	396.0 ± 0.8	82.7 ± 0.3
Z veto	268.3 ± 40.9	837.1 ± 21.2	379.6 ± 5.5	85.9 ± 0.6	67.6 ± 0.3
$N_{\mathrm{j}} \geq 2$	18.9 ± 10.8	475.9 ± 16.1	121.6 ± 3.3	74.7 ± 0.6	52.3 ± 0.3
$p_{\mathrm{T}}^{miss} \geq 50$ GeV	-	346.2 ± 13.6	74.0 ± 2.5	54.9 ± 0.6	42.3 ± 0.3

Table 6.3: Evolution of the number of expected events for important background processes when subsequently applying requirements for the off-Z baseline selection. The number of events corresponds to an integrated luminosity of 35.9 fb^{-1}. The shown uncertainty reflects the statistical uncertainty arising from the finite number of simulated events.

Process	DY	$\mathrm{t\bar{t}}$	WZ	$\mathrm{t\bar{t}Z}$	$\mathrm{t\bar{t}W}$
$\geq$ 3 leptons, $p_{\mathrm{T}} > 10$ GeV	3120.5 ± 142.2	1107.3 ± 24.2	2912.4 ± 15.8	396.8 ± 0.8	82.7 ± 0.3
$m_{\ell\ell} \geq 12$ GeV	3120.5 ± 142.2	1107.3 ± 24.2	2912.4 ± 15.8	396.8 ± 0.8	82.7 ± 0.3
lepton p_{T} selection	3087.4 ± 141.6	1092.6 ± 24.2	2893.4 ± 15.8	396.0 ± 0.8	82.7 ± 0.3
Z veto selection	2819.4 ± 135.5	255.5 ± 11.6	2513.8 ± 14.7	310.1 ± 0.8	14.7 ± 0.1
$N_{\mathrm{j}} \geq 2$	220.7 ± 37.5	149.7 ± 9.1	868.6 ± 8.6	286.0 ± 0.8	11.4 ± 0.0
$p_{\mathrm{T}}^{miss} \geq 50$ GeV	51.8 ± 16.9	105.5 ± 7.7	524.6 ± 6.6	198.8 ± 0.6	9.2 ± 0.0
$p_{\mathrm{T}}^{miss} \geq 70$ GeV (low N_{b}, p_{T}^{miss} and H_{T})	8.3 ± 5.8	82.7 ± 6.9	376.3 ± 5.8	167.9 ± 0.5	8.1 ± 0.0

Table 6.4: Evolution of the number of expected events for important background processes when subsequently applying requirements for the on-Z baseline selection. The number of events corresponds to an integrated luminosity of 35.9 fb^{-1}. The shown uncertainty reflects the statistical uncertainty arising from the finite number of simulated events.

In order to estimate the signal selection efficiency and to assess the sensitivity of the analysis, different mass splitting scenarios of the T1tttt and the T5qqqqWZ topologies have been simulated as benchmark models. For each model, one uncompressed and one more compressed spectrum is considered in order to probe both extreme scenarios. In the case of T1tttt, both scenarios feature a gluino mass of 1200 GeV, while the LSP mass is set to 100 GeV and 700 GeV, respectively. For the T5qqqqWZ model the compressed (uncompressed) scenario is represented with a gluino mass of 800(1000) GeV and a LSP

mass of 500(100) GeV. Table 6.5 shows that a large fraction of the simulated signal events passes the baseline selection. The selection efficiency for events with at least three leptons is larger than 71% for the 2 different mass scenarios of the T1tttt model and larger than 83% for the 2 scenarios of the T5qqqqWZ model. It can also be seen that the increased p_{T}^{miss} requirement for on-Z events with small b jet multiplicities, low H_{T}, and low p_{T}^{miss} does not affect the signal efficiency significantly.

	off-Z		on-Z	
$m_{\tilde{g}}/m_{\tilde{\chi}_1^0}$ (GeV)	T1tttt	T1tttt	T5qqqqWZ	T5qqqqWZ
selection	(1200/100)	(1200/700)	(1000/100)	(800/500)
≥ 3 leptons, $p_{\mathrm{T}} > 10$ GeV	76.3 ± 2.2	80.6 ± 3.0	63.2 ± 2.6	234.2 ± 11.0
$m_{\ell\ell} \geq 12$ GeV	76.2 ± 2.2	80.0 ± 3.0	63.2 ± 2.6	233.3 ± 11.0
lepton p_{T} selection	76.2 ± 2.2	80.0 ± 3.0	63.2 ± 2.6	232.9 ± 10.9
Z veto (Selection)	61.2 ± 2.0	59.8 ± 2.8	57.1 ± 2.5	219.4 ± 10.6
$N_{\mathrm{j}} \geq 2$	61.0 ± 2.0	59.5 ± 2.8	51.1 ± 2.5	213.6 ± 10.5
$p_{\mathrm{T}}^{miss} \geq 50$ GeV	60.0 ± 2.0	57.3 ± 2.8	55.9 ± 2.5	205.0 ± 10.3
$p_{\mathrm{T}}^{miss} \geq 70$ GeV (low N_{b}, p_{T}^{miss} and H_{T})	60.0 ± 2.0	57.3 ± 2.8	55.9 ± 2.5	195.0 ± 10.3

Table 6.5: Evolution of the number of expected events for selected mass points of the T1tttt and T5qqqqWZ SUSY benchmark model. The yields are normalized to an integrated luminosity of 35.9 fb^{-1}. The shown uncertainty reflects the statistical uncertainty arising from the finite number of simulated events.

Figures 6.6 and 6.7 show the distributions of important event observables - H_{T}, p_{T}^{miss}, N_{b} and the transverse mass reconstructed with a lepton and the missing transverse momentum vector,

$$M_{\mathrm{T}} = \sqrt{2p_{\mathrm{T}}^{\ell} p_{\mathrm{T}}^{miss} \left[1 - \cos\left(\phi_{\ell} - \phi_{\vec{p}_{\mathrm{T}}^{miss}}\right)\right]},$$

where one can see the signal to background discrimination power of these variables. If the event is categorized as on-Z, the M_{T} is calculated with the lepton that is not involved in the Z boson mass reconstruction, otherwise, the lepton yielding the lowest M_{T} value is used in the computation of this variable. For background processes containing a W boson, the M_{T} has a well defined end point, while for signal this end point is lost, even if there is a W boson, due to the large p_{T}^{miss} produced by the LSP. The number of events is normalized to 35.9 fb^{-1}. For the off-Z selection, Fig. 6.6 shows that the dominant background arises from nonprompt leptons and that the shapes of the distributions are distinctively different for the SM background processes compared to the T1tttt signal model. The latter exhibits significantly larger b jet multiplicities as well as larger p_{T}^{miss}, H_{T}, and M_{T} values. Additionally, it can be observed that the uncompressed spectrum produces events with larger values of p_{T}^{miss} and H_{T} owing to the larger amount of energy available in the decay chain. Similar behavior can be observed in Fig. 6.7, where the SM background is compared to the two benchmark scenarios of the T5qqqqWZ model in the on-Z baseline region. Depending on the mass splitting between the gluino and the LSP, different shapes of H_{T} and p_{T}^{miss} are expected, which would lead to different levels of an excess of data over the background prediction in the tails of these distributions. Owing to the production of light flavor only in the T5qqqqWZ model, the distribution of the b jet multiplicity does not show a distinctively different shape for signal and background for this model.

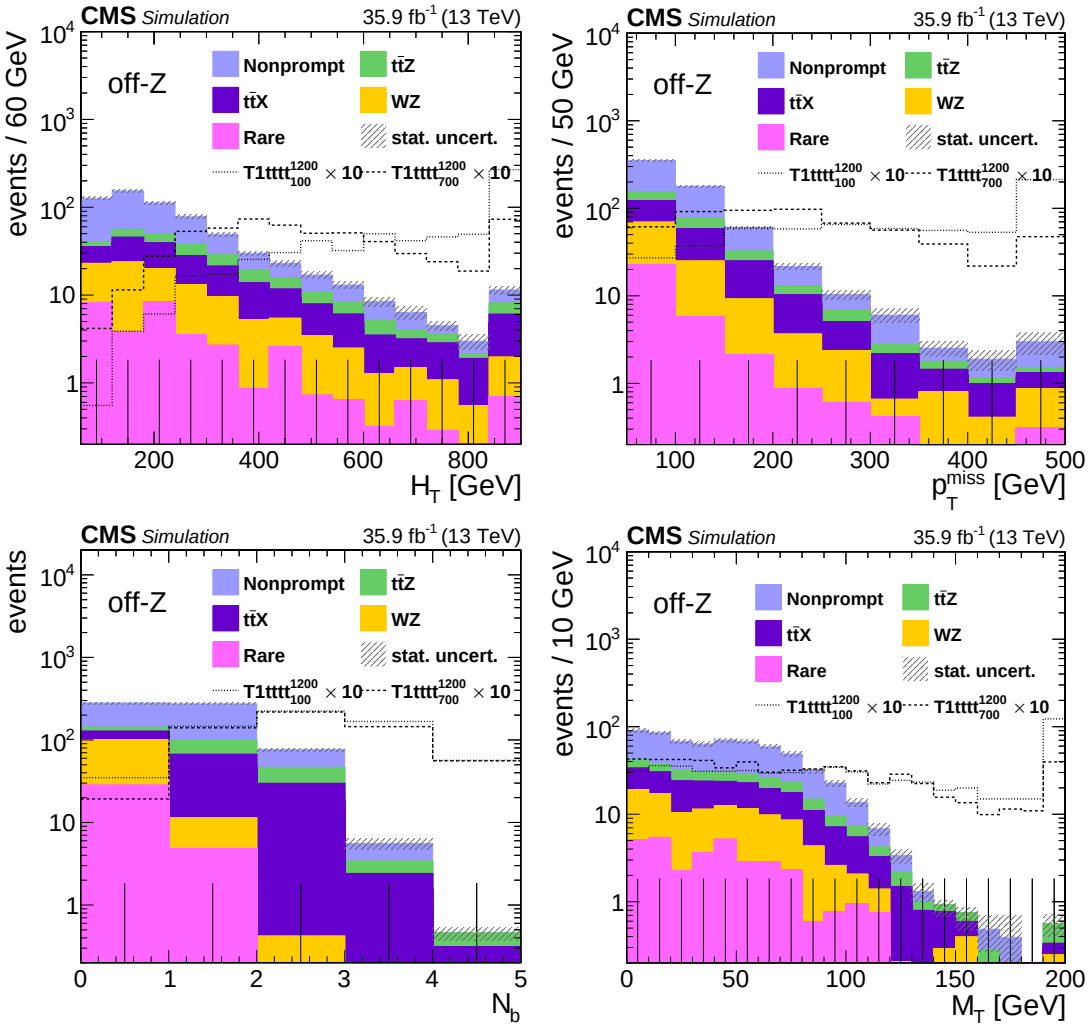


Figure 6.6: Distributions of H_{T}, p_{T}^{miss}, N_{b}, and M_{T} for background processes and selected SUSY benchmark models after applying the off-Z baseline selection. The hatched bands show the statistical uncertainty arising from the finite number of simulated events.

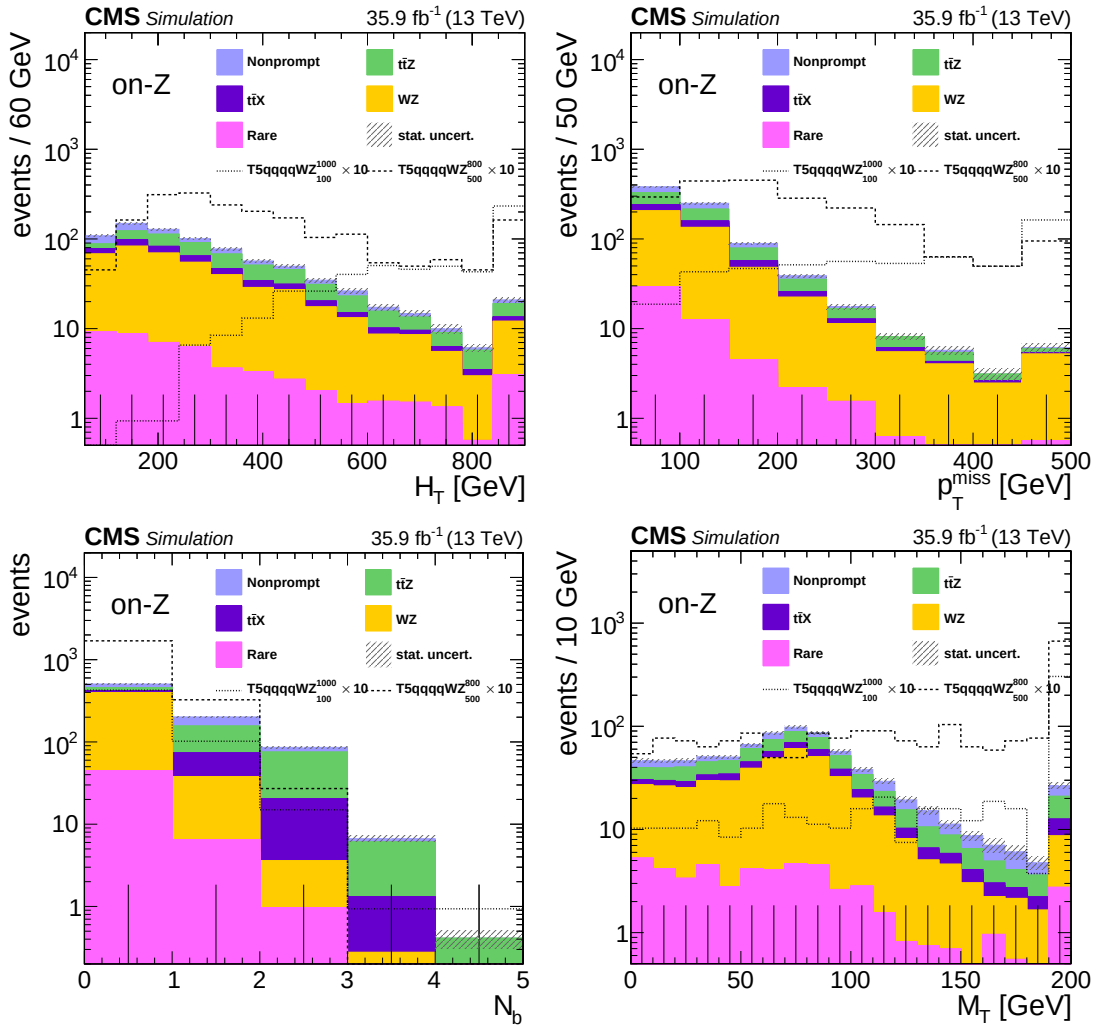


Figure 6.7: Distributions of H_{T}, p_{T}^{miss}, N_{b}, and M_{T} for background processes and selected SUSY benchmark models after applying the on-Z baseline selection. The hatched bands show the statistical uncertainty arising from the finite number of simulated events.

6.3.2 *Signal regions*

To improve further the sensitivity of the search, events that pass the baseline selection are categorized into different exclusive SRs to enhance the signal-to-background ratio for specific signal topologies. The SRs are defined separately for off-Z and on-Z events by a set of boundary conditions on the event observables H_T, p_T^{miss}, N_b and M_T. The categorization in N_b maximizes the sensitivity for signal models that yield multiple top quarks such as the T1tttt model. Such models populate SRs with large b jet multiplicities while important background processes, e.g., WZ and nonprompt lepton backgrounds are mainly contained in SRs requiring low N_b. Nevertheless, also background dominated SRs with low b jet multiplicities are analyzed for this search. Firstly, they are used to gain confidence in the predicted background yields and furthermore these analysis categories help constraining the uncertainties on the backgrounds that populate this region. Secondly, such SRs allow probing signal models that do not produce top quarks like the T5qqqqWZ model. The sensitivity for this topology, which predominantly populates on-Z SRs, profits from the fact that the contribution from the nonprompt lepton background is mainly contained in the off-Z regions. Intermediate b jet multiplicity categories are expected to be useful for T6ttWW and T6ttHZ, where we expect on average 2 b quarks in the final state.

Additional event categorization according to the observed p_T^{miss}, H_T, and M_T allows to further enhance the signal-to-background ratio as more stringent requirements on these variables increasingly suppress SM background. Moreover, such categorization separates compressed and uncompressed signal models. The SR boundaries have been optimized to separate different background sources and to maximize sensitivity for the SUSY benchmark models.

All SRs are summarized in Table 6.6. There are 16 separate on-Z and 16 off-Z SRs. For each category four regions are defined for events with 0, 1, and 2 b jets, depending on the amount of H_T and p_T^{miss}. These SRs are denoted as SR 1-12. Motivated by the low expected yield of events with high b jet multiplicities, one inclusive SR with $p_T^{miss} < 300$ GeV and $H_T < 600$ GeV has been defined for ≥ 3 b jets (SR 13). Additionally to this SR, three extra SRs with significant amounts of H_T (> 600 GeV, SRs 14, 15) or p_T^{miss} (> 300 GeV, SR 16) have been introduced. These latter three regions are inclusive in N_b. All of the 0 b jet regions, as well as three regions with high H_T and p_T^{miss} values, are further split depending on whether M_T is smaller (designated with the letter "a" after the region number) or greater (designated with "b") than 120 GeV, leading to a total of 23 regions for each of the on-Z and off-Z categories.

N_j	N_b	H_T (GeV)	50(70) GeV $\leq p_T^{miss} <$ 150 GeV	150 GeV $\leq p_T^{miss} <$ 300 GeV	$p_T^{miss} \geq$ 300 GeV
≥ 2	0	60 – 400	SR1 †	SR2 †	SR16 †
		400 – 600	SR3 †	SR4 †	
	1	60 – 400	SR5	SR6	
		400 – 600	SR7	SR8	
	2	60 – 400	SR9	SR10	
		400 – 600	SR11	SR12	
	≥ 3	60 – 600	SR13		
	inclusive	≥ 600	SR14 †	SR15 †	

Table 6.6: Multilepton SR definition, used both for on-Z and off-Z events. Regions marked with a † are further split according to whether M_T is smaller or larger than 120 GeV. $p_T^{miss} \geq 70$ GeV is only required in on-Z SR1 and SR5 regions.

6.4 BACKGROUND ESTIMATION

The analysis strategy mentioned in the previous section relies on comparing the observed data and the yields from known SM processes in each SR category defined in the previous section. This means that a robust and precise estimation of the different background contributions is essential to obtain a reliable result. This section describes the methods used for estimating SM background contributions and their validation in data and simulation.

The SM processes that produce less than three prompt leptons that can enter the selection if at least one nonprompt lepton passes all identification and isolation requirements. We commonly refer to this type of background as nonprompt lepton background. The estimation of this background is achieved using data samples that are orthogonal to the signal enriched data samples. This procedure is referred to as "data-driven" in LHC language and more details on this will be given in the following subsection.

Background from WZ and $t\bar{t}Z$ production is the dominant background contribution in various on-Z SRs. The background source is estimated using MC simulation and a simultaneous fit in an orthogonal to the baseline region control regions enriched in these two processes are used to obtain the normalization for these processes.

Other SM background sources are estimated from MC simulation. The most important contribution amongst the remaining background sources are the processes containing top quark(s) in association with a W, Z or BEH boson or another pair of top quarks.

6.4.1 *Nonprompt lepton background*

The main source of the nonprompt lepton background in this analysis, as it could be seen from the Tables 6.3 and 6.4, is the production of $t\bar{t}$ in association with jets. In this process two prompt leptons can originate from leptonic decays of the W bosons that are produced in the top quark decays and the contribution to p_{T}^{miss} arises from neutrinos. Events with these two prompt leptons can enter the baseline selection if an additional nonprompt lepton, i.e. a lepton from heavy flavor hadron decay, misidentified hadrons, muons from light meson decays, or electrons from unidentified photon conversions, passes the tight lepton selection. Even though the tight selection suppresses the vast majority of the $t\bar{t}$ background, the cross section of this process is high enough to ensure $t\bar{t}$ production to be the main source of background in most of the off-Z SRs.

The amount of nonprompt lepton background depends on the probability for a nonprompt lepton to pass the tight selection, i.e. a parameter that depends on the performance of the detector to identify and reconstruct nonprompt leptons. The simulation of this background is rather complex and less reliable than the simulation of processes producing prompt leptons. Therefore, an estimation technique that is independent of MC simulations is used to predict the contribution from nonprompt lepton background. This technique is referred to as a tight-to-loose method.

Since discrimination of prompt and nonprompt leptons that pass the tight selection is essentially impossible, the method predicts the contribution of the nonprompt lepton background in a given SR by weighting the number of events measured in an orthogonal sideband region with the tight-to-loose ratio (TLR). This region is populated by multilepton events for which at least one lepton does not pass the tight lepton selection, but satisfies the loose criterion. This region is called an application region. All other kinematical criteria are identical to the ones used in a particular SR. The tight-to-loose ratio depends on the misidentification rate, f, defined as the probability of a loosely

identified lepton to satisfy the tight selection criterion. In case of n leptons passing loose identification criterion and failing tight selection, the TLR is equal to

$$\mathrm{TLR} = (-1)^{n+1} \prod_{i=1}^{n} \frac{f_i}{1 - f_i}, \tag{6.1}$$

where the product is running over all such leptons [119].

The misidentification rate is a key component in the nonprompt lepton background estimation. It strongly depends on the flavor and on the kinematic properties of the lepton, and thus it needs to be determined as a function of the lepton p_{T} and $|\eta|$ individually for electrons and muons. The measurement of the misidentification rate is performed in the QCD multijet events, which is enriched in nonprompt leptons. Suitable events are preselected by a set of auxiliary single lepton triggers listed in Table B.2. The selection of events entering the denominator of the misidentification rate calculation requires exactly one lepton that passes the loose selection, while in the numerator we require this lepton to pass the tight identification criterion. Electroweak processes, involving W and Z bosons and producing a prompt lepton, in the selection are suppressed by additionally requiring one recoiling jet with $\Delta\mathrm{R}(\ell, \mathrm{jet}) > 1.0$, $p_{\mathrm{T}}^{jet} > 30$ GeV, and p_{T}^{miss} and M_{T} to be less than 20 GeV. The remaining contamination from electroweak processes is subtracted using estimates from the simulation. The normalization of the simulation used for this subtraction is corrected by applying electroweak scale factors, which are derived from a data to MC comparison in a control region enriched in electroweak processes. This control region is defined by requiring p_{T}^{miss} to be greater than 20 GeV, 70 GeV $< M_{\mathrm{T}} <$ 120 GeV, and exactly one lepton to pass the tight lepton selection.

The tight-to-loose method assumes that the misidentification rate measured in the QCD multijet events is the same as a misidentification rate exhibited in the application regions. To mitigate the dependence of the tight-to-loose ratio on the flavor composition of the jets from which the nonprompt leptons originate, the tight-to-loose ratio is parameterized as a function of a variable that correlates more strongly with the parent parton p_{T} than with the lepton p_{T}. Technically it is done by correcting the lepton p_{T} with the energy around the lepton. For the leptons that pass the requirement $p_{\mathrm{T}}^{\mathrm{rel}} > p_{\mathrm{T}}^{\mathrm{rel,\ cut}}$, the transformation for p_{T} is given by

$$p_{\mathrm{T}} \to p_{\mathrm{T}}(1 + \max(0, I_{\mathrm{mini}} - I_{\mathrm{mini}}^{\mathrm{cut}})), \tag{6.2}$$

while the p_{T} of the leptons for which the above requirement does not take place is transformed as

$$p_{\mathrm{T}} \to \max(p_{\mathrm{T}}, p_{\mathrm{T}}^{jet} \times p_{\mathrm{T}}^{\mathrm{ratio,\ cut}}).$$

The values for $I_{\mathrm{mini}}^{\mathrm{cut}}$ and $p_{\mathrm{T}}^{\mathrm{ratio,\ cut}}$ are defined in Table 5.2. The transformed p_{T}, also referred to as cone-corrected p_{T}, serves as a proxy for the parent parton p_{T}. Note that the transformation only modifies the p_{T} of leptons that pass loose selection and fail tight selection and does not affect leptons in events that enter the SRs. The flavor dependence, which is much more important for the case of electrons, is further reduced by relaxing the electron MVA value to obtain similar values for nonprompt electrons that originate from light- or heavy-flavor jets. A similar matching is not needed in the case of muons since nonprompt leptons of this flavor almost exclusively originate from b jets.

The typical misidentification rate both for electrons and muons is around 10% in the corrected lepton p_T between 15 and 45 GeV. For low-p_T muons with 10 GeV $< p_T <$ 15 GeV the misidentification rate increases significantly to about 50%. Low-p_T electrons exhibit a smaller misidentification rate around 30%. For high p_T leptons with $p_T >$ 45 GeV the tight-to-loose ratio is increased to approximately 15-20%.

VALIDATION IN SIMULATION AND DATA The validation of the nonprompt lepton background estimation is performed in a so-called closure test. The closure test shows the capability of the tight-to-loose method to correctly predict the nonprompt lepton background. This is done by comparing the number of events directly observed in the baseline region with the number of predicted events, where the prediction is based on the number of events observed in the corresponding application region reweighted by the tight-to-loose ratio. This test is fully based on simulation, therefore, the misidentification rates used for the prediction have been measured in simulated QCD multijet events using the same prescription as for the measurement of the misidentification rate in data.

The closure test with a $t\bar{t}$ MC sample as a function of main variables used in the search is shown in Figure 6.9 and 6.8 for on-Z and off-Z regions, respectively. In order to maximize the statistical precision, both muons and electrons are selected for the closure test. To verify that the closure works for both flavors independently, the agreement between predicted and observed yields is checked in events with different flavor composition.

The estimation of the nonprompt lepton background is additionally verified with data enriched in nonprompt lepton background. This data sample is formed by selecting three leptons that pass all nominal identification, isolation, and p_T requirements in order to obtain a region similar to the SRs of this analysis. Selecting events with 30 GeV $< p_T^{miss} <$ 50 GeV ensures orthogonality to the baseline selection. The control region is enriched with nonprompt lepton background by applying an off-Z selection for the three leptons and by requiring N_j to be equal to 1 or 2 and $N_b \geq 1$. With this selection, a purity of 80% in the nonprompt lepton background is achieved. A comparison of the observed data and the nonprompt lepton background as predicted by the tight-to-loose method is shown in Figure 6.10 for the full data set of 35.9 fb^{-1} of proton-proton collisions. The small contributions from other SM background sources are taken from MC simulations and the uncertainty band includes both the statistical and systematic uncertainties, the latter of which will be discussed in detail in section 6.5. The signal contamination in this validation region was found to be negligible for the benchmark signals introduced in Section 6.1.

6.4.2 *WZ and ttZ background*

The WZ process is one of the main backgrounds in the regions with 0 b jets, while $t\bar{t}Z$ gives a significant contribution in categories enriched in b jets. The estimation of these backgrounds is based on MADGRAPH5_aMC@NLO MC simulations, but the normalizations are obtained from a simultaneous fit using two control regions enriched in one of the processes. The validation region for WZ production is defined requiring three leptons to pass tight selection among which two leptons have to form an OCSF pair with invariant mass 15 GeV close to the mass of a Z boson. A number of jets and b jets has to be ≤ 1 and 0, respectively. The p_T^{miss} has to be in the range between 30 and 100 GeV, and the invariant mass of a third lepton not participating in the forming of a

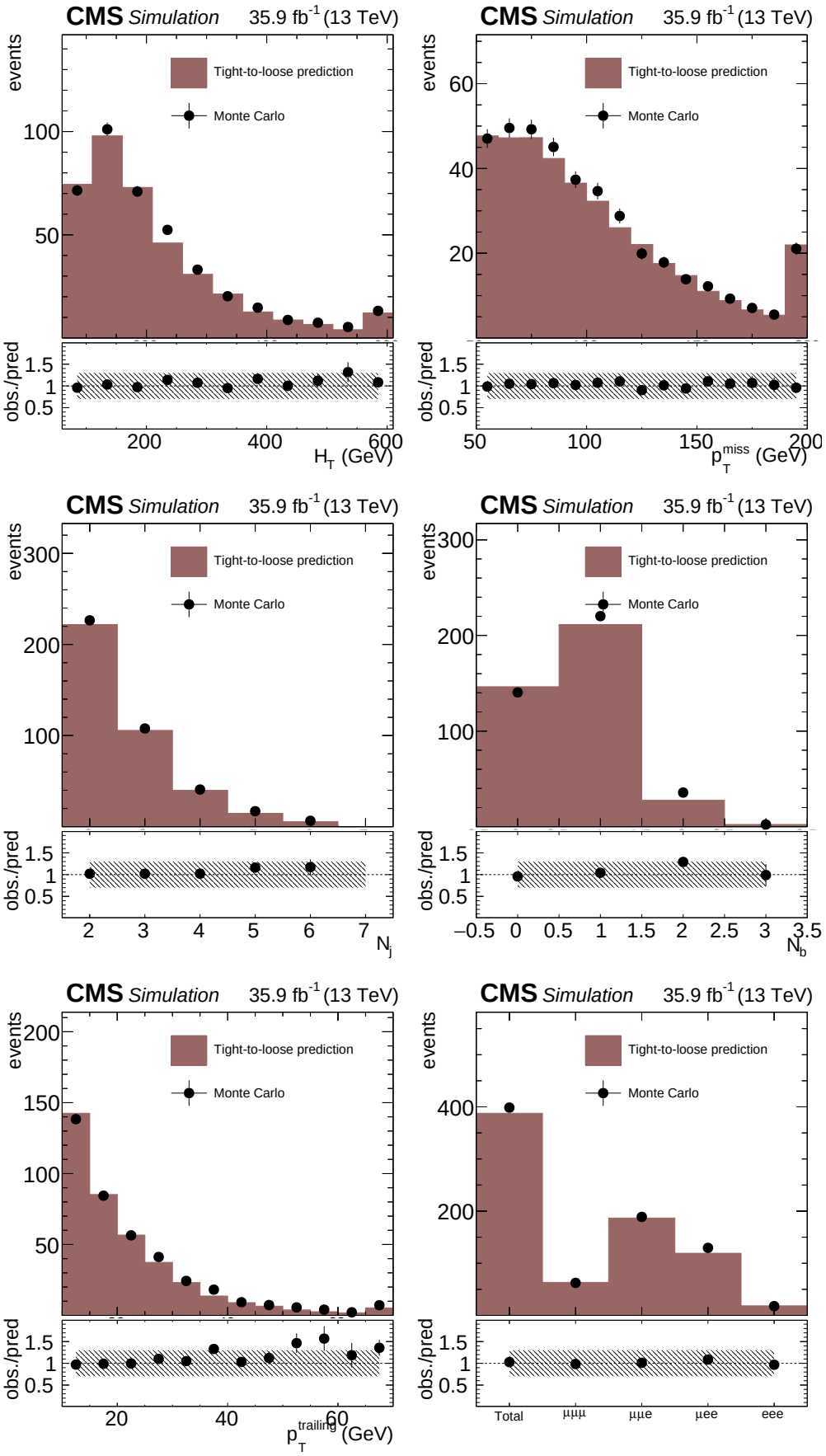


Figure 6.8: Closure test for the nonprompt lepton background estimation in $t\bar{t}$ simulated events in the off-Z region. The misidentification rate is measured in a QCD multijet MC sample. The distributions of H_T, p_T^{miss}, N_j, N_b, and p_T of the trailing lepton are shown. Additionally, the yields in events with different lepton flavor composition are shown.

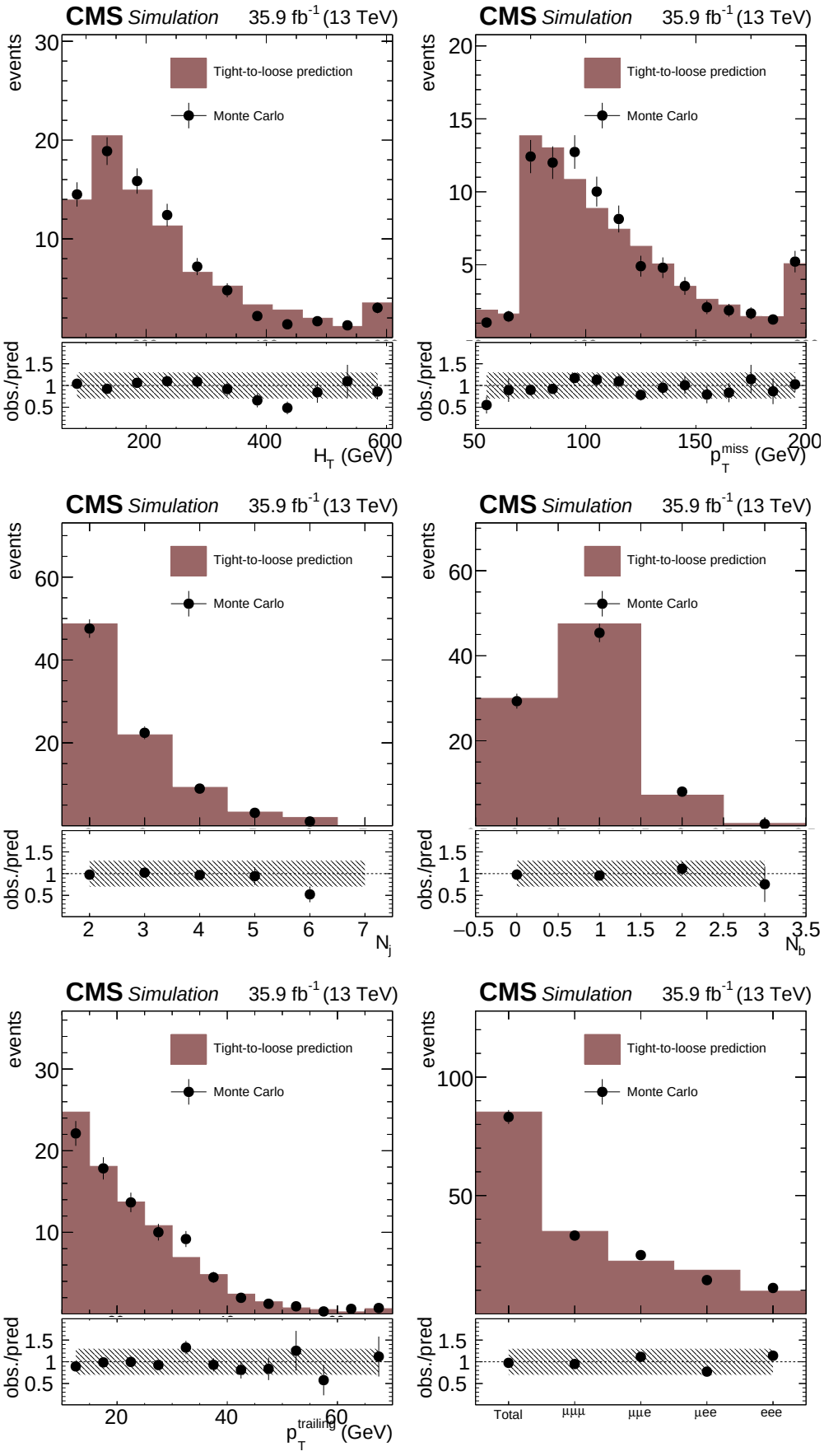


Figure 6.9: Closure test for the nonprompt lepton background estimation in $t\bar{t}$ simulated events in the on-Z region. The misidentification rate is measured in a QCD multijet MC sample. The distributions of H_T, p_T^{miss}, N_j, N_b, and p_T of the trailing lepton are shown. Additionally, the yields in events with different lepton flavor composition is shown.

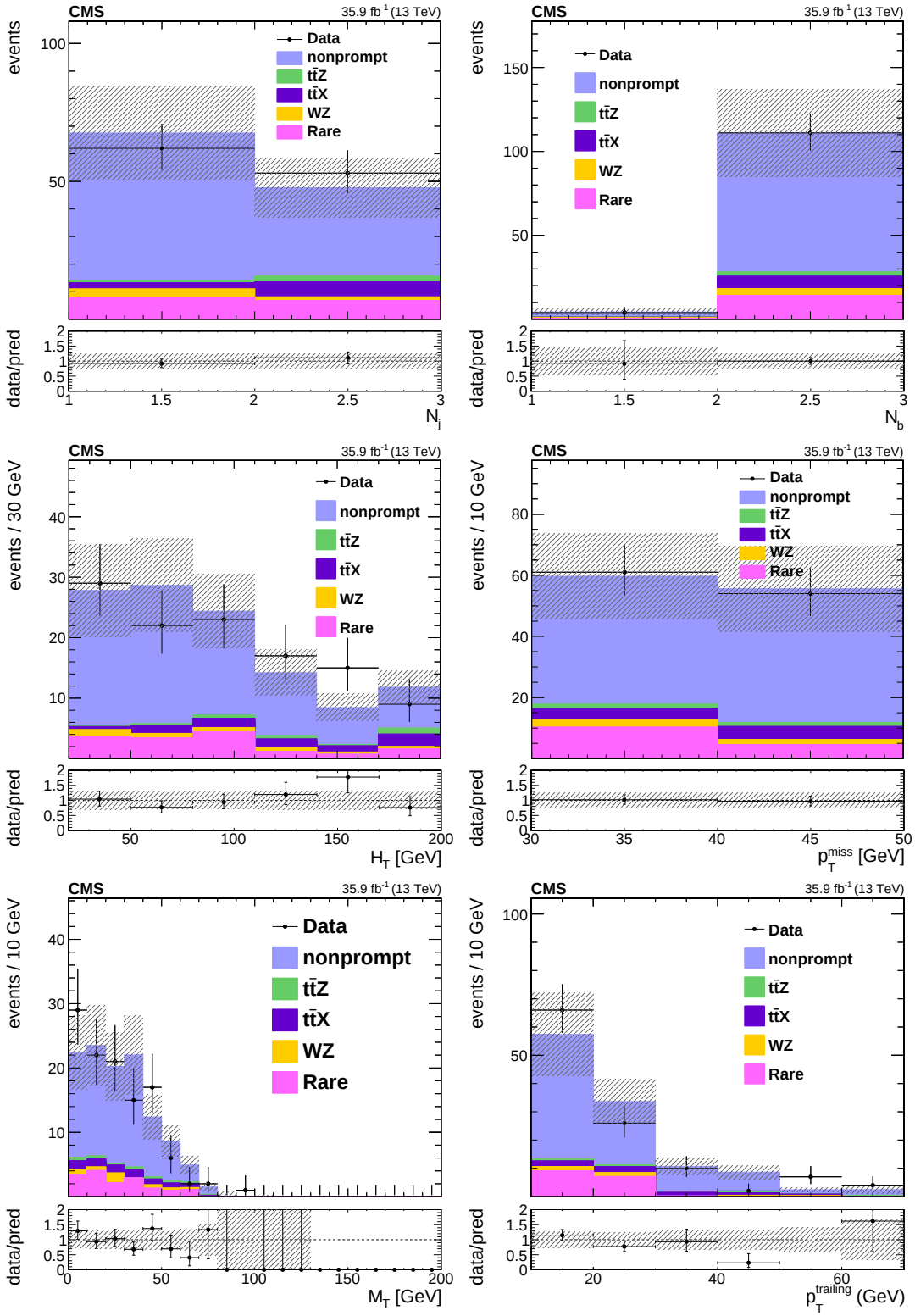


Figure 6.10: The data control region enriched in nonprompt lepton background is shown. Good agreement between data and nonprompt lepton background estimated with a tight-to-loose method is observed in N_{j}, N_{b}, H_{T}, p_{T}^{miss}, M_{T} and the trailing lepton p_{T} distributions for an integrated luminosity of 35.9 fb^{-1}

.

Z boson is required to be at least 50 GeV in order to suppress contamination from the DY process. The purity of such a region is 80% in WZ, which is shown in Figure 6.11.

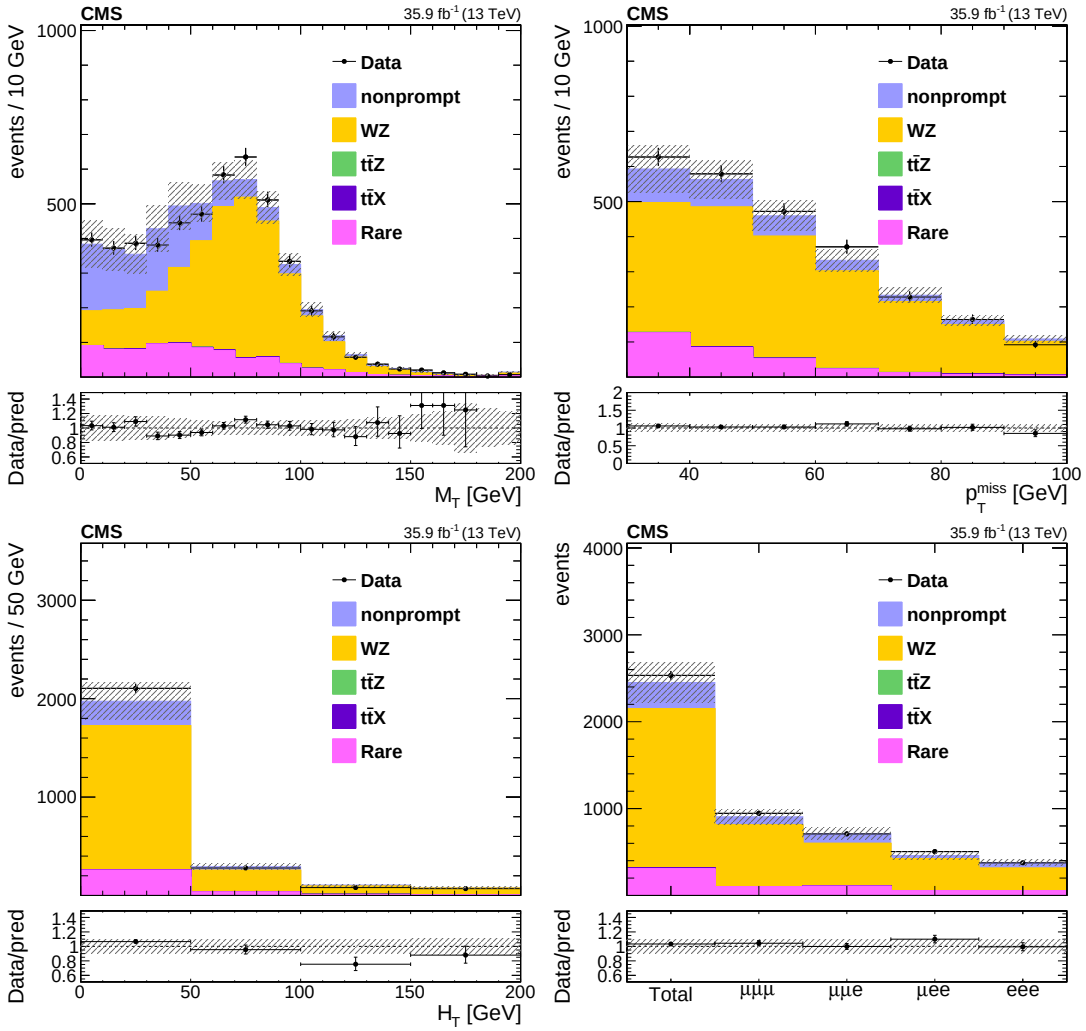


Figure 6.11: The control region enriched in the WZ production is shown. The nonprompt lepton background is estimated with a tight-to-loose method and other backgrounds are estimated with a MC simulation. For M_{T} plot the requirement on $M_{\mathrm{T}} > 50$ GeV is removed to show the contribution from nonprompt leptons in the $N_{\mathrm{b}} = 0$ category. Additionally, the distributions for p_{T}^{miss}, H_{T} and flavor composition are shown after applying all selection criteria described in the text. The hatched bands represent the statistical and systematic uncertainties.

The region enriched in $\mathrm{t\bar{t}Z}$ production is defined in the same way as for WZ, apart from a requirement on the number of jets: three tight leptons, one OCSF pair 15 GeV close to the mass of a Z boson, $N_{\mathrm{j}} \geq 3$, and $30 < p_{\mathrm{T}}^{miss} < 50$ GeV. Events are classified by the number of b jets, so three categories are formed for the $\mathrm{t\bar{t}Z}$ CR: $N_{\mathrm{b}} = 0$, 1, and ≥ 1. The overall purity of the $\mathrm{t\bar{t}Z}$ process is approximately 20% and increases to 50% in the bins with at least one b jet. The $\mathrm{t\bar{t}Z}$ control region is shown in Figure 6.12.

These three bins, together with the WZ region, are used in a simultaneous fit to obtain the scale factors for the normalization of the simulated samples. In the fit to data, the normalization and relative population across all four bins of all the components are allowed to vary according to experimental and theoretical uncertainties, which will be

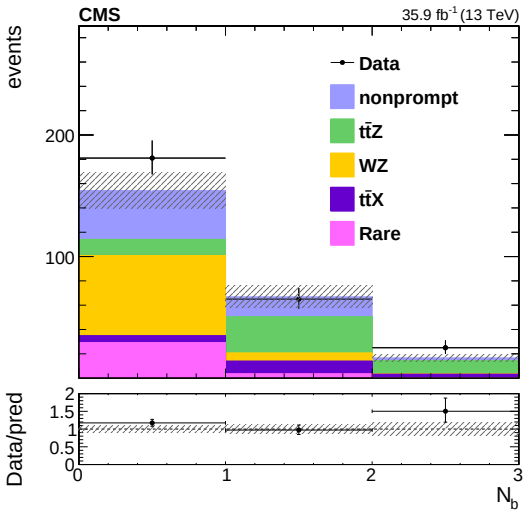


Figure 6.12: The control region enriched in the $t\bar{t}Z$ production is shown. The nonprompt lepton background is estimated with a tight-to-loose method and other backgrounds are estimated with a MC simulation. The hatched bands represent the statistical and systematic uncertainties.

discussed in Section 6.5. For the WZ process, the obtained scale factor is compatible with unity, 1.01 ± 0.07, and no correction is applied to the simulation, while for the $t\bar{t}Z$ it is found to be 1.14 ± 0.28. Therefore the yields from the MC $t\bar{t}Z$ sample obtained in the baseline region are scaled by a factor of 1.14. Based on this fit, the uncertainty in the normalization of the WZ process is estimated to be 10%, while a value of 25% uncertainty is assigned for the $t\bar{t}Z$ production.

6.4.3 *Other background processes*

Several SM processes can produce events that enter the baseline selection and give the same signature as SUSY processes. All processes in this category are characterized by having a production rate similar to the SUSY particles. Specifically, the main contribution originates from the production of a single or a pair of top quarks in association with a vector boson or BEH boson or another pair of top quarks. The majority of these processes are simulated with MADGRAPH5_aMC@NLO generator, with some exception like for the $t\bar{t}H$ process, which is produced with the POWHEG generator (for details see Table B.3). The contribution from these processes is estimated using the output from the MC simulations, scaled to the luminosity available in data and on top of this estimation the scale factors described in Section 5.3 are applied. All contributions from processes in this category are denoted as $t\bar{t}X$.

Processes that contain a virtual photon decaying into leptons are referred to as internal conversions. The processes in this category must not be confused with external conversions where a real photon converts into leptons upon interaction with the detector material. All processes in the internal conversion group, namely $t\bar{t}\gamma$, $Z\gamma$, $t\gamma$ and $W\gamma$, are simulated at NLO accuracy with the MadGraph5_aMC@NLO event generator. The internal conversion background components are strongly suppressed by the baseline selection, specifically requirements on p_{T}^{miss} and N_{j}. The background originating from external conversions is largely suppressed by the requirement for tracks to have hits in all pixel detector layers and its estimate is included in the nonprompt lepton background.

Among the remaining background processes, the largest background is ZZ diboson production, which can yield up to 4 leptons in the final state and contributes to on-Z low b jet categories. It is simulated with the POWHEG event generator. This generator is also used to model the contribution from BEH boson production through gluon fusion with subsequent leptonic decay via two Z bosons ($gg \to H \to ZZ$). The remaining processes in this category are simulated with MadGraph5_aMC@NLO and include tri-boson production and the production of a BEH boson in association with a vector boson. These processes and internal conversion background processes are grouped together in a rare SM backgrounds category.

6.5 SYSTEMATIC UNCERTAINTIES

Systematic uncertainties affecting the estimated yields of signal and background processes for the various SRs can bias the result of this search. All systematic uncertainties can be categorized as experimental, e.g. uncertainties related to the JEC or the b tagging efficiency; as theoretical, e.g. the uncertainties on the considered cross sections; and as statistical owing to the limited size of the MC samples. Moreover, dedicated uncertainties on the nonprompt lepton background estimated from data are considered.

The impact of the systematic uncertainties is evaluated by varying an estimated yield in the signal or background category under study up and down by the extent of the systematic uncertainty. This variation can either affect similarly the yields in all SRs (e.g. for the uncertainty on the luminosity) or fluctuate from one SR category to another (e.g. for JEC and b tagging uncertainties). A summary of the magnitudes of the systematic uncertainties in this search and the resulting effect on the yields in signal and background estimates are given in Table 6.7 and described in more detail below.

Source	Magnitude	Effect on yield	Nonprompt	WZ, $t\bar{t}Z$	MC bkg.	signal
luminosity	2.6%	2.6%			✓	✓
JEC	1 – 8% per jet	1 – 10%		✓	✓	✓
b tag efficiency	1 – 10% per jet	1 – 10%		✓	✓	✓
pileup	4.6%	1 – 5%		✓	✓	✓
lepton efficiencies	3% per ℓ	9%			✓	✓
HLT efficiencies	3%	3%			✓	✓
lepton eff. FastSim	2% per ℓ	6%				✓
appl. region stat.	10 – 100%	10 – 100%	✓			
non-prompt extrapol.	30%	30%	✓			
EWK subtraction (fit)	100% (ewk. SF)	5%	✓			
WZ CR normalization	10%	10%		✓		
ttZ CR normalization	25%	25%		✓		
QCD scales	×0.5 / ×2	11 – 13 % (σ) 3 – 18% ($\mathcal{A}$)		✓($t\bar{t}Z$, $\mathcal{A}$)	$t\bar{t}W$, $t\bar{t}H$	
PDFs	–	2 – 3%			✓	
Rare SM bkgs.	50%	50%			Rare SM processes	
ISR modeling	1 – 10%	1 – 10%				✓
Modeling of unclustered energy	1 – 20%	1 – 20%				✓
MonteCarlo stat.	1 – 100%	1 – 100%		✓	✓	✓

Table 6.7: Summary of the sources of uncertainties, their magnitude and the associated effect on the yield of selected events. The second column indicates the magnitude of a given uncertainty and the third column shows the effect on the number of expected background and signal events when varying up and down the uncertainty by 1 standard deviation. The last four columns indicate which processes are affected by the respective uncertainty, where 'MC bkg.' stands for $t\bar{t}X$ and rare SM processes.

6.5.1 *Experimental uncertainties*

The first experimental uncertainty is the imperfect knowledge of JEC. This type of uncertainty affects all backgrounds and signals taken from the simulation. The jet energy scale is varied by 1-8%, depending on p_{T} and η of the jet. The impact of this uncertainty is assessed by shifting the jet energy correction factors for each jet up and down by one standard deviation before the calculation of all kinematic quantities. The JEC uncertainties are propagated to the p_{T}^{miss} and all jet- and p_{T}^{miss}-related variables (N_{j}, H_{T}, N_{b}, and M_{T}) used in this analysis. In signal regions which are not statistically limited, variations of 1-10% of the yields are observed when the JEC is varied by one standard deviation.

A similar approach is used for the uncertainties associated with the corrections for the b tagging efficiencies for light and b flavor jets, which are parametrized as a function of p_{T} and η. The variation of the b tagging correction factor is at maximum of the order of 10% per jet, which leads to an overall effect on the yield of up to 20% depending on the SR.

All simulation yields are reweighted to match the PU distribution measured in data, based on the number of vertexes. The inelastic proton-proton cross section of 69.2 mb [63] is known within 4.6%, so the variations of the minimum bias cross section within this uncertainty are propagated to the reweighting and through the full chain of the analysis to estimate the associated uncertainty on the yields. The observed effect on the yield is of the order of 1-5%. Additional uncertainty is taken into account to model the precision of the luminosity measurement of CMS. Except for the WZ and $\mathrm{t\bar{t}Z}$ processes, all background and signal processes that are estimated from MC simulations and scaled to the luminosity available in data, are affected by this uncertainty. For the data set of 35.9 fb^{-1} analyzed in this thesis, an uncertainty of 2.6% is quoted on the luminosity measurement.

For the lepton reconstruction and identification, the corrections applied to simulation as explained in Section 5.3 have their own uncertainties. A flat uncertainty of 3% per muon is considered, while a p_{T} and η dependent uncertainty of a similar magnitude is considered for electrons. An additional 1% systematic uncertainty is applied to electrons with $p_{\mathrm{T}} < 20$ or > 80 GeV. The uncertainties for different leptons are conservatively assumed to be correlated, leading to about 9% uncertainty for events with three leptons. In contrast to the uncertainty of the b tagging efficiency scale factors, this uncertainty can be propagated linearly to the number of predicted background events. The uncertainty related to the HLT trigger efficiency correction for the simulated backgrounds amounts to $\pm 3\%$. For events with a correction factor larger than 0.97 the up variation of this uncertainty naturally does not exceed an efficiency of 100%.

All correction factors and associated uncertainties have also been measured for the signal processes generated with a "FastSim" simulation. Correction factors for b tagging and lepton reconstruction, identification and isolation are applied on top of the data to the "FullSim" correction and come with dedicated uncertainties. The application of the various corrections and the treatment of their uncertainties follows the same procedure as for the corrections for the full simulation described above.

6.5.2 *Uncertainties related to data-driven background estimations*

Three different uncorrelated uncertainties are considered for the estimation of the nonprompt lepton background. The first is a SR independent uncertainty of 30% on the

prediction, where the magnitude is motivated by the level of agreement between the observed and the predicted number of simulated events in the closure test presented in Subsection 6.4.1.

A second contribution accounts for the uncertainty associated with the simulation-based subtraction of electroweak processes in the nonprompt lepton enriched region used for the misidentification rate measurement. The measured correction factors found in the electroweak control region are varied within the uncertainty, which results in two alternative misidentification rates for each flavor parametrized in the nominal cone-corrected p_{T} and $|\eta|$ binning. Propagating the alternative misidentification rates through the full background estimation procedure allows deriving SR dependent upward and downward variations of the estimated number of background events with respect to the nominal value.

The statistical power of the application region represents the third source of uncertainty of the final estimate for the nonprompt lepton background. This uncertainty follows directly from the number of events entering the application region and varies from 10 to 100% depending on the SR. A dedicated approach is employed in SRs with missing events in the application regions. In the full 2016 data set, this only concerns 2 on-Z application regions (SR 3b and 15b) and 1 off-Z region (14b). The upper bound of the expected event number from nonprompt lepton background category in these regions is estimated by multiplying the upper uncertainty of a zero event expectation from a Poisson distribution by a most likely misidentification rate.

6.5.3 *Theoretical uncertainties*

The effect of the theoretical uncertainties on the cross section and on the acceptance have been studied separately for $\mathrm{t\bar{t}W}$ and $\mathrm{t\bar{t}H}$. In order to assess the effect of the uncertainty of the renormalization and factorization scales, all possible permutations of independent upward and downward variations of the scales have been considered. For the upward variation, the nominal value for both scales has been doubled, while it has been decreased by half for the downward variation. The largest effect has been found when both scales are varied in the same direction and the associated effect on the cross section was found to be about 13% for $\mathrm{t\bar{t}W}$ and 11% for $\mathrm{t\bar{t}H}$ background, respectively. For another important background - $\mathrm{t\bar{t}Z}$- a 25% flat uncertainty from normalization in the control region is assigned on the cross section. The effect of the upward and downward variations on the acceptance has been also studied by comparing the resulting differences in the number of background events in the different SR with respect to the nominal prediction. The effect is taken into account by considering an additional uncorrelated uncertainty of 3–18%, depending on the SR and on the regarded process.

The uncertainty related to the knowledge of PDF is estimated from a set of 100 variations of the NNPDF3.0 PDFs [120]. The uncertainty on the cross section and the acceptance effect is considered together [121] and estimated as the root-mean-square of the deviations with respect to the nominal number of background events. Since no significant fluctuations among the different signal regions are seen, a flat uncertainty of 3% (2%) is applied for $\mathrm{t\bar{t}W}$ ($\mathrm{t\bar{t}H}$).

The same theoretical uncertainties also affect the $\mathrm{X}\gamma$ and the remaining rare SM background processes. An uncertainty of 50% is considered for the contribution from these processes to take into account all of the effects described above. Due to the very small number of expected events from these processes, this uncertainty has only a small impact on the sensitivity of the search.

For the signal dedicated uncertainties related to the calculation of p_{T}^{miss} in the "FastSim" package and to the uncertainty on the ISR modeling are considered. The uncertainty in potential mismodeling of p_{T}^{miss} in the "FastSim" arises from unclustered energy in the CMS detector and is evaluated by comparing the reconstructed p_{T}^{miss} with the p_{T}^{miss} obtained using generator-level information. The mean of the signal contribution in the two cases is taken as the central value and half of the difference between the two values is assigned as systematic uncertainty, correlated among all SRs.

Uncertainties on the production cross section of the signal models are taken into account separately as uncertainty bands in the exclusion limit plots.

6.5.4 *Uncertainties arising from limited Monte Carlo statistics*

The limited size of the generated MC samples causes an additional source of uncertainty. For the backgrounds that are estimated from the simulation, like WZ, $\mathrm{t\bar{t}Z}$, $\mathrm{t\bar{t}X}$, and rare SM processes as well as for all the signal processes, this uncertainty is computed from the number and the weights of MC events entering the signal regions. The size of this effect depends on the considered process and signal region and varies between 1 and 100%. The yields and statistical uncertainties for each type of background estimated from MC are shown in Appendix B.3.

6.6 RESULTS

Comparisons between data and predicted background in four event observables used for signal region categorization, namely H_{T}, p_{T}^{miss}, M_{T} and N_{b}, as well as the lepton p_{T} spectra, the lepton flavor composition, and the event jet multiplicity are shown in Figures 6.14 and 6.13 for events satisfying the selection criteria of the on-Z and off-Z, respectively. Figure 6.15 graphically represents a summary of the predicted background and observed event yields in the individual SR bins. The same information is also presented in Tables 6.8 and 6.9.

A good agreement between data and the prediction is observed in all sensitive to analysis variables, with the exception of one large deviation in the three last bins in the leading p_{T} distribution in the off-Z region and in the two last bins of H_{T} distribution in the on-Z region. In the off-Z region where the $p_{\mathrm{T}} > 320$ GeV selection was applied, 18 events were observed with total estimated background of 9.9 ± 1.4. Among these 18 events, 11 of them were observed in the 0 b-tagged category, 14 events in the 0 or 1 jets category and 14 events with p_{T}^{miss} less than 150 GeV. In On-Z category the discrepancy in H_{T} greater than 780 GeV region was observed. In total 38 events were recorded with high hadronic level, while the expected background estimate gives us 24.9 ± 3.4 events. Most of these events contribute to the low b-tagged and low p_{T}^{miss} categories - 29 and 20 events, respectively. No accumulation of the events in any other distribution has been observed, therefore these events are considered as a statistical fluctuation. Besides these two excesses, in the bulk and the tails of all other distributions the predicted total SM background agrees with the observed data within its uncertainty.

The presented results can be interpreted in terms of upper limits on the production cross sections of simplified SUSY signal topologies. This allows to set exclusion limits on SUSY particle masses within the considered signal models described in Chapter 6.1. In the absence of significant deviations between the observed data and the SM background expectation, such exclusion limits allow to compare the sensitivity of the search with previous or similar competing searches that are sensitive to the same signal topologies.

N_b	H_T [GeV]	p_T^{miss} [GeV]	M_T [GeV]	Expected [events]	Observed [events]	SR
0	60 – 400	50 – 150	< 120	$206 \pm 6 \pm 35$	201	SR1a
			≥ 120	$1.4 \pm 0.5 \pm 0.2$	3	SR1b
		150 – 300	< 120	$25.9 \pm 2.1 \pm 4.3$	24	SR2a
			≥ 120	$0.84 \pm 0.34 \pm 0.12$	0	SR2b
	400 – 600	50 – 150	< 120	$15.6 \pm 1.6 \pm 2.1$	21	SR3a
			≥ 120	$0.19 \pm 0.09 \pm 0.02$	0	SR3b
		150 – 300	< 120	$6.0 \pm 0.8 \pm 0.7$	5	SR4a
			≥ 120	$0.19 \pm 0.09 \pm 0.04$	0	SR4b
1	60 – 400	50 – 150	Inclusive	$202 \pm 6 \pm 44$	191	SR5
		150 – 300		$25.6 \pm 1.9 \pm 4.6$	25	SR6
	400 – 600	50 – 150		$15.4 \pm 1.3 \pm 2.2$	21	SR7
		150 – 300		$7.3 \pm 1 \pm 1.1$	7	SR8
2	60 – 400	50 – 150	Inclusive	$47.7 \pm 2.8 \pm 7.6$	51	SR9
		150 – 300		$5.3 \pm 0.5 \pm 0.6$	5	SR10
	400 – 600	50 – 150		$5.8 \pm 0.7 \pm 0.8$	9	SR11
		150 – 300		$2.9 \pm 0.5 \pm 0.4$	2	SR12
≥ 3	60 – 600	50 – 300	Inclusive	$3.9 \pm 0.7 \pm 0.6$	6	SR13
Inclusive	≥ 600	50 – 150	< 120	$14.4 \pm 1.2 \pm 1.6$	20	SR14a
			≥ 120	$0.28 \pm 0.14 \pm 0.04$	0	SR14b
		150 – 300	< 120	$12.1 \pm 1.4 \pm 1.6$	10	SR15a
			≥ 120	$0.40 \pm 0.12 \pm 0.05$	0	SR15b
	≥ 60	≥ 300	< 120	$12.1 \pm 1.5 \pm 1.9$	7	SR16a
			≥ 120	$0.70 \pm 0.25 \pm 0.11$	0	SR16b

Table 6.8: Expected and observed yields in the off-Z search regions. The first uncertainty states the statistical uncertainty, while the second represents the systematic uncertainty.

Additionally, exclusion limits are important measures to reject theoretically viable SUSY models that are disfavored in observed data. Limits on SUSY particle masses in simplified models, however, do not rule out the existence of such SUSY particles if they are part of more complex signal topologies.

6.6.1 *Statistical analysis*

The result of this search for SUSY processes of squark or gluino pair production can be interpreted in terms of exclusion limits on the masses of the SUSY particles involved in the decay cascades in the SMS topologies. The following statistical procedure, known as modified frequentist CL_s method, is used to set the limits and is summarized in [122].

The procedure is based on comparing the observed number of events in each SR or bin i, with the predicted number of SM background events b_i and the expected number of signal events s_i for a given signal model and fixed SUSY particle masses. The signal cross section can be scaled using a real, positive number μ, the so-called signal strength

N_b	H_T [GeV]	p_T^{miss} [GeV]	M_T [GeV]	Expected [events]	Observed [events]	SR
0	$60-400$	$50-150$	<120	$266 \pm 5 \pm 39$	241	SR1a
			≥ 120	$30 \pm 2 \pm 4$	33	SR1b
		$150-300$	<120	$53.8 \pm 2.2 \pm 8$	61	SR2a
			≥ 120	$5.7 \pm 0.8 \pm 0.7$	9	SR2b
	$400-600$	$50-150$	<120	$44.6 \pm 1.9 \pm 6.5$	52	SR3a
			≥ 120	$5.1 \pm 0.6 \pm 0.7$	6	SR3b
		$150-300$	<120	$16.6 \pm 1.3 \pm 2.5$	17	SR4a
			≥ 120	$1.43 \pm 0.33 \pm 0.2$	1	SR4b
1	$60-400$	$50-150$	Inclusive	$116 \pm 4 \pm 15$	115	SR5
		$150-300$		$21.7 \pm 1.2 \pm 2.8$	19	SR6
	$400-600$	$50-150$		$25.2 \pm 1.2 \pm 3.6$	25	SR7
		$150-300$		$7.5 \pm 0.8 \pm 1$	9	SR8
2	$60-400$	$50-150$	Inclusive	$47 \pm 1.6 \pm 7.4$	64	SR9
		$150-300$		$7.2 \pm 0.8 \pm 1.2$	6	SR10
	$400-600$	$50-150$		$11.7 \pm 1 \pm 2.1$	12	SR11
		$150-300$		$2.6 \pm 0.4 \pm 0.4$	6	SR12
≥ 3	$60-600$	$50-300$	Inclusive	$4.7 \pm 0.5 \pm 0.9$	5	SR13
Inclusive	≥ 600	$50-150$	<120	$33 \pm 2 \pm 4$	42	SR14a
			≥ 120	$4.6 \pm 0.6 \pm 0.6$	6	SR14b
		$150-300$	<120	$15.8 \pm 1.2 \pm 2$	13	SR15a
			≥ 120	$1.9 \pm 0.3 \pm 0.2$	4	SR15b
	≥ 60	≥ 300	<120	$19.1 \pm 1.1 \pm 2.8$	23	SR16a
			≥ 120	$2.28 \pm 0.35 \pm 0.26$	5	SR16b

Table 6.9: Expected and observed yields in the on-Z search regions. The first uncertainty states the statistical uncertainty, while the second represents the systematic uncertainty.

modifier. The total number of expected events can then be written as a sum of expected background and signal events in particular bin i: $\mu s_i + b_i$. From the latter, one can easily deduce the meaning of the signal strength modifier: different values of μ represent different hypotheses of the signal. For example $\mu = 0$ corresponds to the background-only hypothesis (H_0), while $\mu = 1$ represents a "signal+background" hypothesis (H_1). The probability of given hypothesis to be represented by data can be described by the likelihood function, where the probability of observing n events with expected $\mu s_i + b_i$ obeys the binomial distribution, which in the limit of large numbers of trial can be approximated with a Poisson distribution. The corresponding likelihood for observing the data for some fixed signal strength μ can be written in the following way:

$$\mathcal{L}(data|\mu) = \prod_i \frac{(\mu s_i + b_i)^{n_i}}{n_i!} e^{-(\mu s_i + b_i)}. \tag{6.3}$$

The multiplication is running over all considered SRs. This likelihood needs to be extended as the predicted numbers of signal and background events s_i and b_i tend to

be biased according to the systematic uncertainties as detailed in Section 6.5. Each uncertainty is described by a so-called nuisance parameter θ, which itself is modeled by a probability density function, $\rho(\theta|\tilde{\theta})$ with an a priori estimate $\tilde{\theta}$ of the nuisances[1]. The Likelihood function therefore will be extended accordingly:

$$\mathcal{L}(data|\mu,\theta) = \prod_i \frac{(\mu s_i(\theta) + b_i(\theta))^{n_i}}{n_i!} e^{-(\mu s_i(\theta)+b_i(\theta))} \rho(\theta|\tilde{\theta}). \tag{6.4}$$

For a given observation, this likelihood function can now be maximized in two different ways. Firstly, we can find nuisance parameters $\tilde{\theta}_\mu$ that maximize $\mathcal{L}$ for a fixed signal strength modifier $\mu \geq 0$, and secondly by letting both the nuisance parameters and the signal strength modifier float to find values $\hat{\theta}$ and $\hat{\mu}$ that correspond to the global maximum of $\mathcal{L}$. Here, $\tilde{\mu}$ is restricted to $0 \leq \tilde{\mu} \leq \mu$, where the lower bound is motivated by the assumption that NP processes cannot reduce the SM background and the upper bound ensures that upward fluctuations of data are not interpreted as evidence confirming the signal hypothesis. The upper bound μ is a fixed parameter in the calculation and is incremented until a certain condition is met as will be discussed further.

A test statistics $\tilde{q}_\mu$ is constructed based on the ratio of these two maximized likelihoods and is referred to as the profile likelihood ratio

$$\tilde{q}_\mu = -2 \log \frac{\mathcal{L}(data|\mu, \hat{\theta}_\mu)}{\mathcal{L}(data|\hat{\mu}, \hat{\theta})}, \tag{6.5}$$

where $0 \leq \hat{\mu} \leq \mu$. This test statistics used in comparing the compatibility of the observation with the H_0 hypothesis on the one hand and the H_1 hypothesis on the other hand.

To proceed further, the value of the test statistics $\tilde{q}_\mu^{obs}$ corresponding to the observed data is evaluated for a signal strength modifier value under test. Then, nuisance parameters $\hat{\theta}_0^{obs}$ and $\hat{\theta}_\mu^{obs}$ are derived that maximize the likelihood function (Equation 6.4) for the H_0 and H_1 hypotheses, respectively. These sets of nuisance parameters are then used to generate toy MC pseudo-data and to construct probability density functions $f(\tilde{q}_\mu|\mu, \hat{\theta}_\mu^{obs})$ and $f(\tilde{q}_\mu|0, \hat{\theta}_0^{obs})$, which describe the test statistic $\tilde{q}_\mu$ for the two aforementioned hypotheses. The level of incompatibility with the respective hypothesis is defined by the p-value p_μ and p_b, which are estimated from the probability density functions:

$$\begin{aligned} p_\mu &= P(\tilde{q}_\mu \geq \tilde{q}_\mu^{obs}|H_1) = \int_{\tilde{q}_\mu^{obs}}^{\infty} f(\tilde{q}_\mu|\mu, \hat{\theta}_\mu^{obs}) d\tilde{q}_\mu \\ 1 - p_b &= P(\tilde{q}_\mu > \tilde{q}_\mu^{obs}|H_0) = \int_{\tilde{q}_0^{obs}}^{\infty} f(\tilde{q}_\mu|0, \hat{\theta}_0^{obs}) d\tilde{q}_\mu. \end{aligned}$$

Finally, the quantity CL_s is calculated as the ratio of the two probabilities:

$$CL_s(\mu) = \frac{p_\mu}{1 - p_b}.$$

For claiming the exclusion of a signal model, the $CL_s < 0.05$ is required for $\mu = 1$, which means that the respective signal is excluded at the 95% confidence level (CL), considering its nominal cross section.

1 for the presented analyses, log-normal distributions are used to model so-called flat uncertainties that affect a given background or signal contribution in all SRs by a constant percentage. Shape uncertainties, whose magnitude vary between the SRs, are modeled with Gaussian distributions

Expected exclusion limits for the H_0 hypothesis can be calculated by generating pseudo-data obtained from the probability density function $f(\tilde{q}_\mu|0,\hat{\theta}_0^{obs})$. This pseudo-data can then be treated as if it were real data in order to calculate CL_s and in this way the signal models with $\mu < 1$ at 95% CL can be separated from the ones with $\mu \geq 1$. This methodology is used to optimize the SR selection and configuration. In general we can say that expected and observed exclusion limits indicate the level of agreement between observed yields and expected background prediction in the most sensitive SR for a considered SUSY scenario. Due to a large number of SUSY models and a number of SUSY particle mass configurations the exclusion limits are calculated with the approximate formula [123] that is valid when the number of expected events is sufficiently large. In addition to the exclusion limits, an upper limit on the production cross section ($\mu \times \sigma_s$) can be calculated, where μ is fulfilling the requirement $CL_s < 0.05$ at 95% CL and σ_s is the nominal production cross section for a particular SUSY scenario.

6.6.2 *Simplified model interpretation*

Figs. 6.16, 6.17 and 6.18 show upper limits on the production cross sections of gluino, bottom and top squark pair productions, respectively. In the T1tttt and T5qqqqWZ models, gluinos with masses smaller than 1610 and 1170 GeV are excluded for a massless $\widetilde{\chi}_1^0$. In the T6ttWW model, the bottom squark masses are excluded up to 840 GeV for charginos lighter than 200 GeV. And finally, in a T6ttHZ simplified model, the $\widetilde{t}_2$ mass is excluded up to 720, 780, or 710 GeV for models with an exclusive decay of $\widetilde{t}_2$ to $\widetilde{t}_1$ and BEH boson, an exclusive decay of $\widetilde{t}_2$ to $\widetilde{t}_1$ and a Z boson, or an equally probable mix of those two decays.

Nearly all search regions that are used in this search have shown to be important for probing various phase-spaces in various models. The region with high p_{T}^{miss} and M_{T} values provides the best sensitivity to non-compressed scenarios of the T1tttt. In compressed cases, the values of p_{T}^{miss} are significantly reduced giving a possibility to intermediate p_{T}^{miss} categories contribute notably in exclusion limits. Additionally, the amount of hadronic activity in non-compressed scenarios remains substantial, therefore the $N_{\mathrm{b}} = 2$ and 400 GeV $< H_{\mathrm{T}} <$ 600 GeV category also play a significant role. No significant excess was observed in these regions, therefore, the expected exclusion limits are compatible with the observed ones.

For the T5qqqqWZ model in the non-compressed scenario, the most sensitive region in the on-Z category are the one inclusive in N_{b}, but with a substantial amount of p_{T}^{miss} and high H_{T} and M_{T} values. The insignificant excess in data was observed in these categories, which is reflected on lower observed limits with respect to expected limits. When moving towards compressed scenarios, the most significant contribution originates from the category with $N_{\mathrm{b}} = 0$ and lower values of H_{T} and p_{T}^{miss}. In this category, the number of expected events agrees with the observed data within its uncertainty, therefore the observed and expected limits are at the same level.

The mass of $\widetilde{\chi}_1^0$ in T6ttWW model is fixed to 50 GeV, while this search targets the exclusion of chargino masses up to 650 GeV. Therefore, a significant amount of p_{T}^{miss} is expected as a signature for this model. In the regions with high p_{T}^{miss}, the overall over-prediction is observed, which drives the observed exclusion limit to be higher than expected.

For the T6ttHZ model with $\mathcal{B}(\widetilde{t}_2 \to \widetilde{t}_1 Z) = 0\%$ branching ratio, the limits in the non-compressed scenario are driven by the off-Z high H_{T} regions, while for the compressed case, the region with significantly lower hadronic activity plays a significant role. Overall

the observed event number in these region is compatible with the expected yields. In scenarios with the 50% and 100% branching ratio the on-Z regions with high N_{b} and H_{T} values play a dominant role. As for the T5qqqqWZ model, in this region an overall small excess in the data is observed, which gives a higher expected limits than the observed ones.

6.6.3 *Comparison with other results*

The exclusion limits placed on all SUSY particle masses in this search extend the limits set by previous searches in CMS that examined the multilepton final state at $\sqrt{s} = 8$ [124] and 13 TeV [118]. In the T1tttt model an improvement of approximately 435 and 250 GeV for gluinos and LSP, respectively, is achieved compared to the exclusion limit set in a similar search based on data collected with the CMS detector in 2015, corresponding to an integrated luminosity of 2.3 fb^{-1}. With respect to the same result, the limit on gluino and neutralino masses in the T5qqqqWZ model extends the corresponding limit by about 335 and 180 GeV, respectively. In T6ttWW model production, the limits on bottom squark and the chargino can be extended by 380 GeV for each particle. And finally, for the model featuring top squark production the limits with respect to the previous result at $\sqrt{s} = 8$ TeV for the model in which $\widetilde{\mathrm{t}}_2$ exclusively decay to $\widetilde{\mathrm{t}}_1$ and BEH boson the limits on $\widetilde{\mathrm{t}}_2$ and $\widetilde{\mathrm{t}}_1$ are extended by 160 GeV and 80 GeV, while in the model which features exclusive decay of $\widetilde{\mathrm{t}}_2$ to $\widetilde{\mathrm{t}}_1$ Z boson the corresponding limits are extended by 170 and 100 GeV.

We can also compare the results in this thesis with results from other groups within CMS. An analysis targeting events with same-charge dilepton pairs plus jets and p_{T}^{miss} in the final state, carried out by CMS at $\sqrt{s} = 13$ TeV with the same data set as the analysis presented in this thesis places higher limits on gluino masses in the non-compressed region in the T1tttt simplified model. Results from searches exploiting a fully hadronic final state and a final state with exactly 1 lepton are higher by approximately $200 - 400$ GeV for the compressed scenario(see Fig. 6.19, left). All analyses within CMS outperform the result obtained in this analysis for the T5qqqqWZ model (see Fig. 6.19, right). This can be explained by the low branching ratio of W and Z bosons decay to a multilepton final state. Nevertheless, the limits placed on the T1tttt and the T5qqqqWZ models in the analysis presented in this thesis have to be seen as a complementary result. The analysis presented here is the only one probing T6ttHZ and T6ttWW within CMS.

The ATLAS collaboration has performed searches for SUSY using the same experimental signature with 36.1 fb^{-1} data of proton-proton collisions [126, 127]. Similar to the ones presented in this thesis, they have also observed no excess of data over the expected SM background and placed limits on SUSY particle masses using similar simplified models. The limits are rather comparable to the one obtained by our analysis.

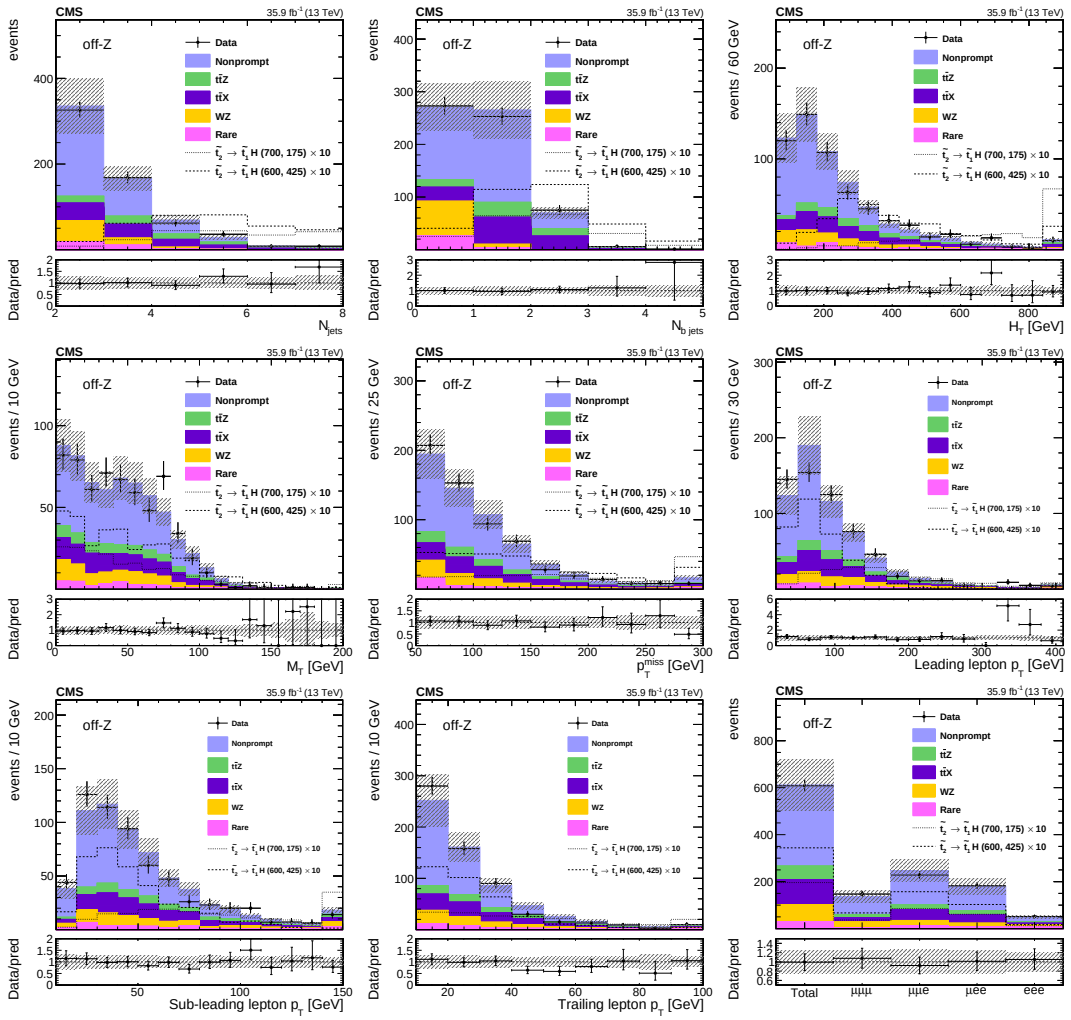


Figure 6.13: Predicted background and observed data yields in the key observables for the off-Z baseline selection: N_{j}, N_{b}, H_{T}, M_{T}, p_{T}^{miss}, the lepton p_{T} spectra and the event yields by flavor category are shown. The last bin includes the overflow events, and the hatched area represents the statistical and combined systematic uncertainties in the prediction. The lower panels show the ratio of the observed and predicted yields in each bin. For illustration the yields, multiplied by a factor 10, for two signal mass points in the T6ttHZ model, where the $\mathcal{B}(\widetilde{\mathrm{t}}_2 \to \widetilde{\mathrm{t}}_1 H) = 100\%$, are displayed for non-compressed ($\mathrm{m}(\widetilde{\mathrm{t}}_2) = 700$ GeV and $\mathrm{m}(\widetilde{\mathrm{t}}_1) = 175$ GeV) and compressed ($\mathrm{m}(\widetilde{\mathrm{t}}_2) = 600$ GeV and $\mathrm{m}(\widetilde{\mathrm{t}}_1) = 425$ GeV) scenarios.

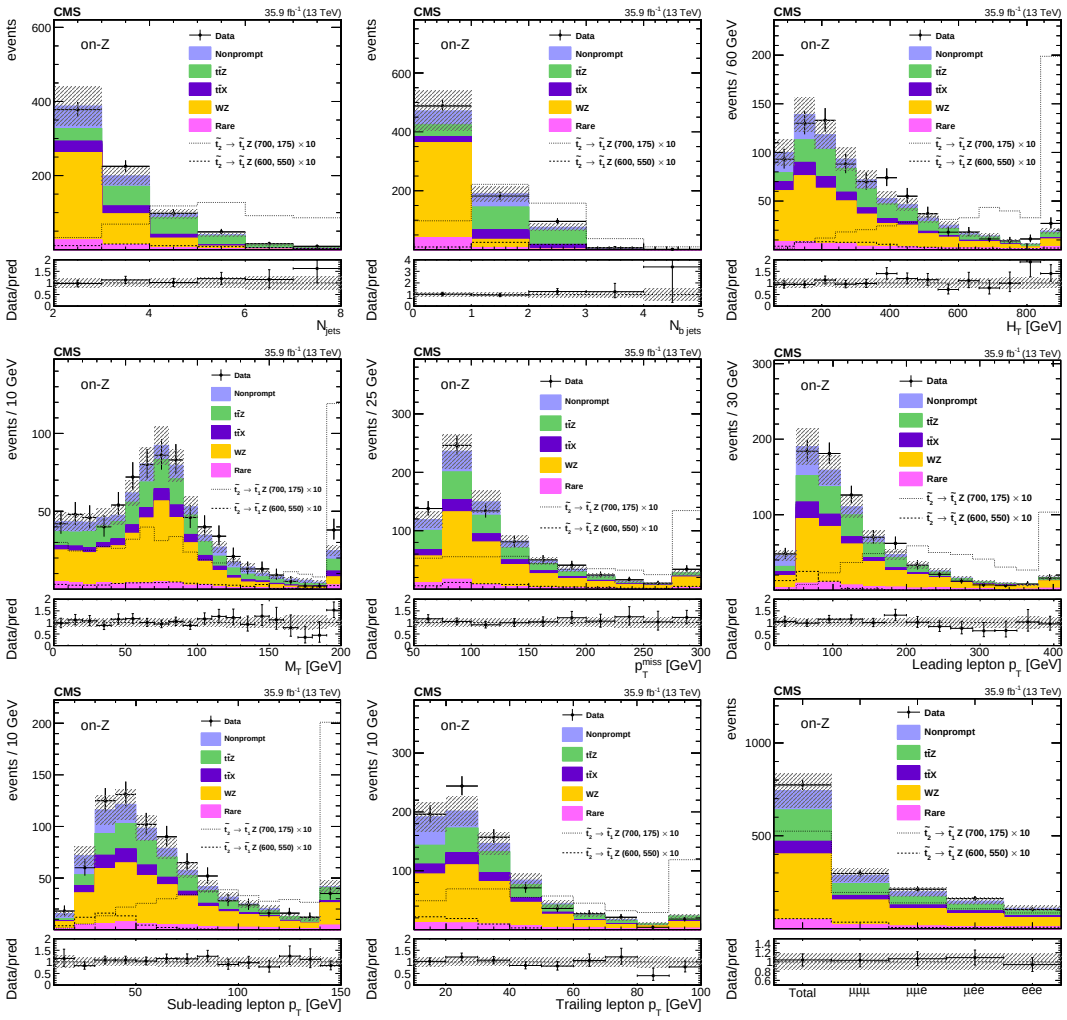


Figure 6.14: Predicted background and observed data yields in the key observables of the on-Z baseline selection: the number of jets and b jets, H_{T}, M_{T}, p_{T}^{miss}, the lepton p_{T} spectra and the event yields by flavor category are shown. The last bin includes the overflow events, and the hatched area represents the combined statistical and systematic uncertainties in the prediction. The lower panels show the ratio of the observed and predicted yields in each bin. For illustration the yields, multiplied by a factor 10, for two signal mass points in the T6ttHZ model, where the $\mathcal{B}(\widetilde{\mathrm{t}}_2 \to \widetilde{\mathrm{t}}_1 Z) = 100\%$, are displayed for non-compressed ($\mathrm{m}(\widetilde{\mathrm{t}}_2) = 700$ GeV and $\mathrm{m}(\widetilde{\mathrm{t}}_1) = 175$ GeV) and compressed ($\mathrm{m}(\widetilde{\mathrm{t}}_2) = 600$ GeV and $\mathrm{m}(\widetilde{\mathrm{t}}_1) = 550$ GeV) scenarios.

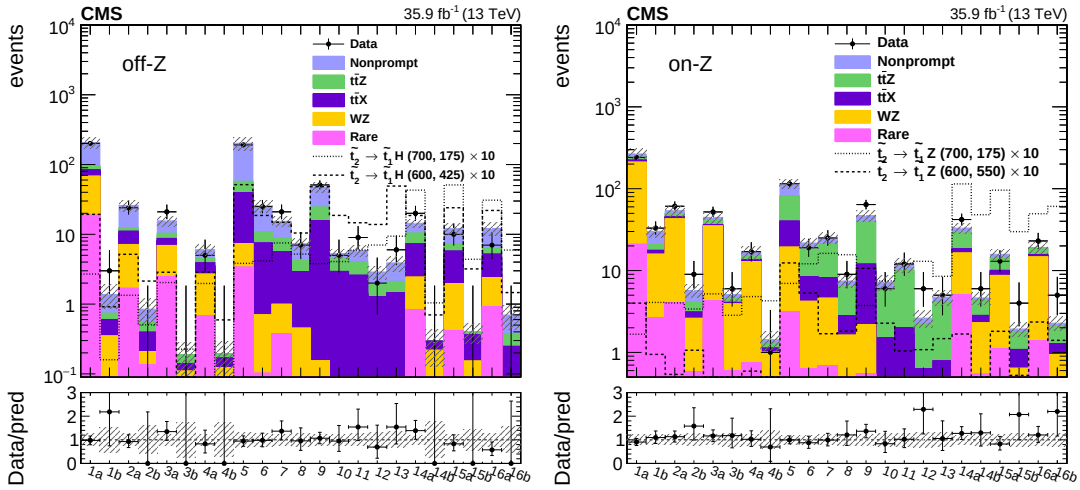


Figure 6.15: Predicted background and observed data yields in the 23 off-Z (left) and the 23 on-Z (right) signal regions. The hatched area represents the statistical and systematic uncertainties on the prediction. The lower panels show the ratio of the observed and predicted yields in each bin. For illustration the yields, multiplied by a factor 10, for $\tilde{t}_2 \to \tilde{t}_1$ H (left) and $\tilde{t}_2 \to \tilde{t}_1$ Z (right) decays are displayed for two signal mass points in the T6ttHZ model to represent compressed and non-compressed scenarios.

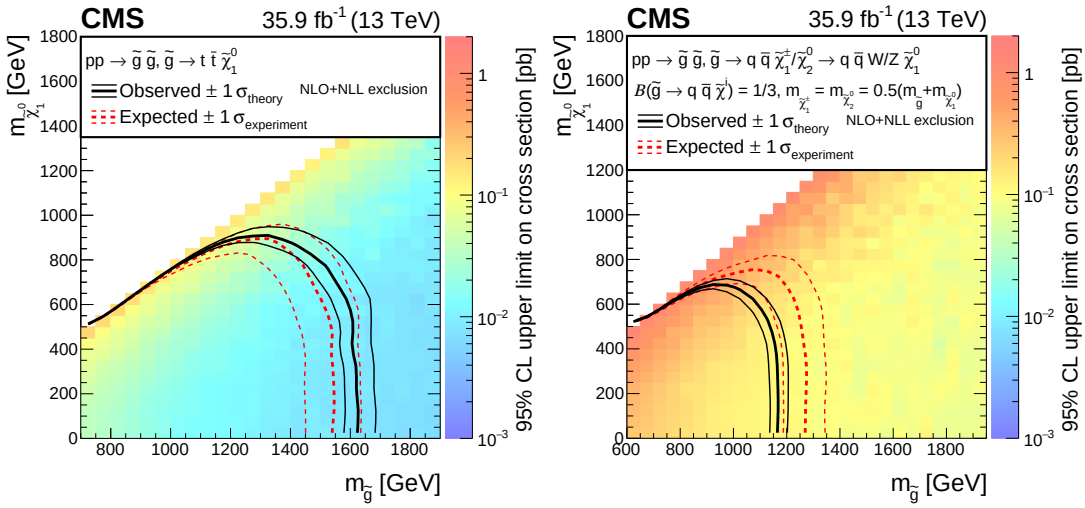


Figure 6.16: Cross section upper limits at 95% confidence level in the $m_{\tilde{\chi}_1^0}$ versus $m_{\tilde{g}}$ plane for T1tttt (left) and T5qqqqVV (right) simplified models. For the latter model the branching fraction of gluino decay to neutralino or chargino is equal to 1/3 and $m_{\tilde{\chi}_1^\pm} = m_{\tilde{\chi}_2^0} = 0.5(m_{\tilde{g}} + m_{\tilde{\chi}_1^0})$. The excluded regions are to the left and below the observed and expected limit curves. The color scale indicates the excluded cross section at a given point in the mass plane.

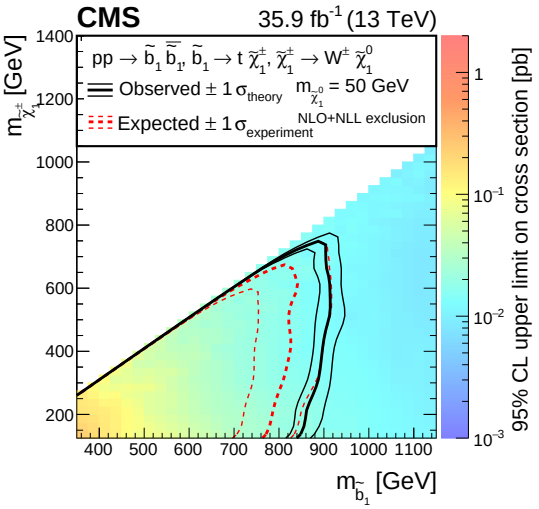


Figure 6.17: Cross section upper limits at 95% CL in the $m_{\widetilde{\chi}_1^\pm}$ versus $m_{\widetilde{b}_1}$ plane for T6ttWW simplified model. The mass of the $\widetilde{\chi}_1^0$ is set to 50 GeV. The descriptions of the excluded regions and color scale are the same as in Fig. 6.16.

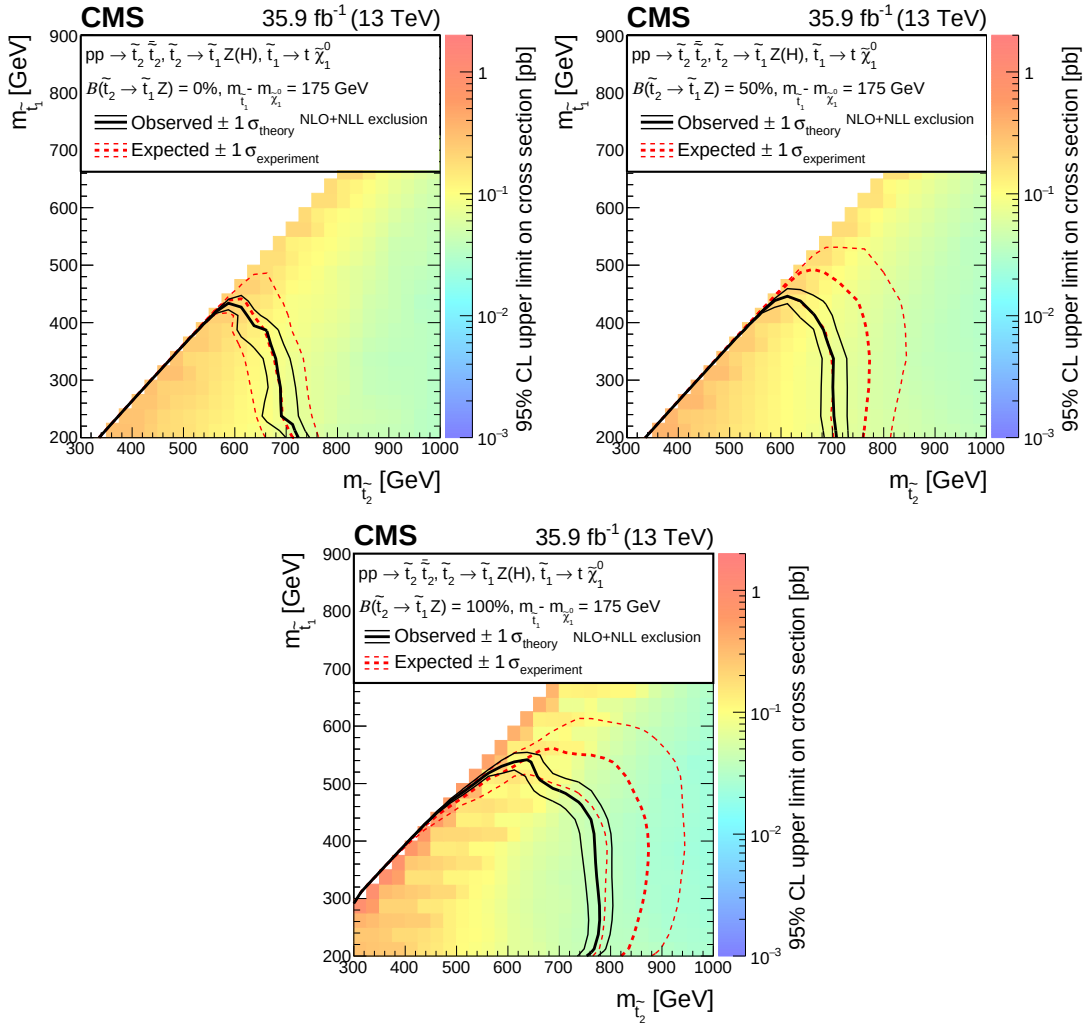


Figure 6.18: Cross section upper limits at 95% CL in the $m_{\tilde{t}_1}$ versus $m_{\tilde{t}_2}$ plane for T6ttHZ simplified model. Different branching fractions of the decay $\tilde{t}_2 \to \tilde{t}_1 Z$ are considered: 0% (top left), 50% (top right), and 100% (bottom). The mass difference between the lighter top squark ($\tilde{t}_1$) and a neutralino is close to the mass of the top quark. The descriptions of the excluded regions and color scale are the same as in Fig. 6.16.

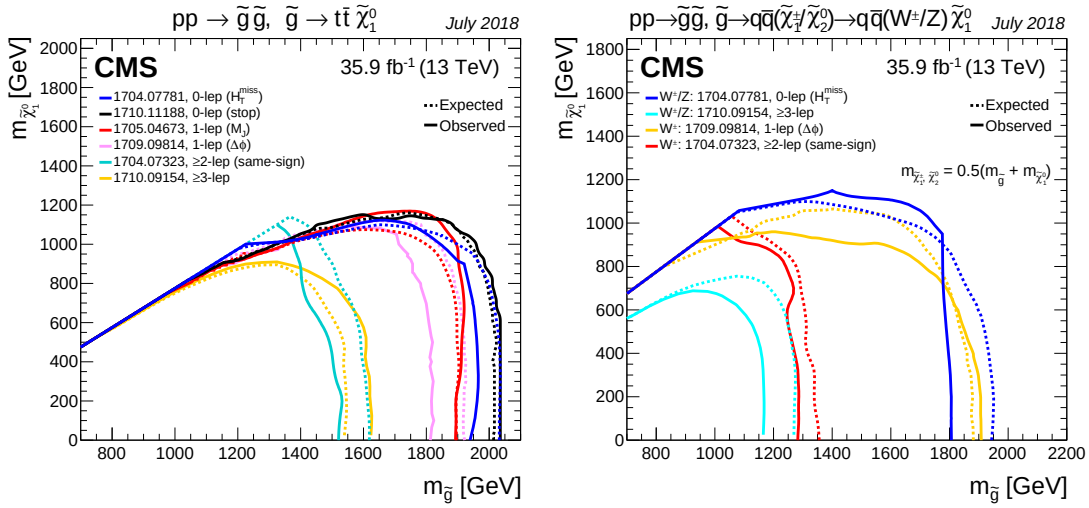


Figure 6.19: Exclusion contours for a simplified model of gluino pair production with gluino decays to pairs of top quarks and the LSP(left) and for a simplified model of gluino pair production with gluino decays via an intermediate chargino ($\tilde{\chi}_1^{\pm}$) or neutralino ($\tilde{\chi}_2^0$) to a final state with a quark, an antiquark, and either a W or a Z boson. The mass of the intermediate gaugino is set to the average of the top squark and LSP masses. The Figures are taken from [125]

Part II

INCLUSIVE AND DIFFERENTIAL CROSS SECTION MEASUREMENT FOR TOP PAIR PRODUCTION IN ASSOCIATION WITH A VECTOR BOSON

7

MEASUREMENT OF THE TOP PAIR PRODUCTION IN ASSOCIATION WITH A W BOSON

In this chapter the measurement of top pair production in association with a W boson, hereafter referred to as $t\bar{t}W$, using the data collected by the CMS detector during 2016 is presented. Experimentally this process was observed for the first time in an analysis of proton-proton collision data at center-of-mass energy of $\sqrt{s} = 8$ TeV by the CMS Collaboration [128]. The data used corresponds to the integrated luminosity of 19.5 fb^{-1} and the measured cross section was 382^{+117}_{-102} fb with a precision level of 30%. The analysis presented in this chapter uses a data set of proton-proton collisions at $\sqrt{s} = 13$ TeV. At such increased center-of-mass energy the $t\bar{t}W$ cross section increases, leading to a larger number of signal events, however, the main background process, $t\bar{t}$, receives an even larger increase in its production cross section. This fact clearly led to challenges and necessitated the usage of involved analysis techniques. The cornerstone of this analysis is the usage of one of such techniques - multivariate analysis which exploits a boosted decision tree classifier with gradient boosting.

The $t\bar{t}W$ process is measured in the most promising decay channel: two leptons with the same charge accompanied by jets and neutrinos (see Fig. 7.1). Requiring the same charge for the two leptons retains only one third of the signal in the dilepton final state. However, this selection significantly improves the signal-to-background ratio, as same charge lepton pairs are produced in the SM processes with relatively small cross sections. The main background in this final state originates from the $t\bar{t}$ process. Having 3 orders of magnitude higher cross section than $t\bar{t}W$, it contributes in two ways to the total background. Firstly, in semileptonic $t\bar{t}$ decay, where only one prompt lepton is available, an additional nonprompt lepton may stem from heavy-quark decay giving rise to the same-charge dilepton yield. Secondly, the charge of one of the two leptons in dileptonic $t\bar{t}$, where both leptons are prompt and come from W, can be misidentified. Data-driven techniques are used to measure and validate these two types of backgrounds.

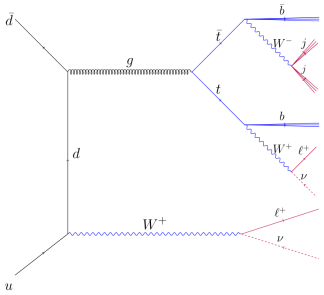


Figure 7.1: The $t\bar{t}W$ process decay in the same-charge dilepton final state. The charged leptons (denoted by ℓ) are accompanied by four jets (j), among which two are b-tagged jets (b), and two neutrinos (ν).

The results on the $t\bar{t}W$ cross section measurement presented in this thesis (together with the $t\bar{t}Z$ measurement and the EFT interpretation of the results) are published by the CMS collaboration. In the following we add the publication as reference. After this, each ingredient of the data analysis will be discussed in detail.

EUROPEAN ORGANIZATION FOR NUCLEAR RESEARCH (CERN)

CERN-EP-2017-286
2018/08/16

CMS-TOP-17-005

Measurement of the cross section for top quark pair production in association with a W or Z boson in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$

The CMS Collaboration*

Abstract

A measurement is performed of the cross section of top quark pair production in association with a W or Z boson using proton-proton collisions at a center-of-mass energy of 13 TeV at the LHC. The data sample corresponds to an integrated luminosity of 35.9 fb^{-1}, collected by the CMS experiment in 2016. The measurement is performed in the same-sign dilepton, three- and four-lepton final states. The production cross sections are measured to be $\sigma(\text{t}\bar{\text{t}}\text{W}) = 0.77^{+0.12}_{-0.11}\,(\text{stat})^{+0.13}_{-0.12}\,(\text{syst})\,\text{pb}$ and $\sigma(\text{t}\bar{\text{t}}\text{Z}) = 0.99^{+0.09}_{-0.08}\,(\text{stat})^{+0.12}_{-0.10}\,(\text{syst})\,\text{pb}$. The expected (observed) signal significance for the $\text{t}\bar{\text{t}}\text{W}$ production in same-sign dilepton channel is found to be 4.5 (5.3) standard deviations, while for the $\text{t}\bar{\text{t}}\text{Z}$ production in three- and four-lepton channels both the expected and the observed significances are found to be in excess of 5 standard deviations. The results are in agreement with the standard model predictions and are used to constrain the Wilson coefficients for eight dimension-six operators describing new interactions that would modify $\text{t}\bar{\text{t}}\text{W}$ and $\text{t}\bar{\text{t}}\text{Z}$ production.

Published in the Journal of High Energy Physics as doi:10.1007/JHEP08(2018)011.

*See Appendix A for the list of collaboration members

1 Introduction

The 13 TeV center-of-mass energy of proton-proton (pp) collisions at the LHC opens the possibility for studying the processes at larger mass scales than previously explored in the laboratory. The top quark-antiquark pair ($t\bar{t}$) produced in association with a W ($t\bar{t}W$) or Z ($t\bar{t}Z$) boson is among the most massive signatures that can be studied with high precision. The theoretical cross sections at next-to-leading order (NLO) in quantum chromodynamics (QCD) for $t\bar{t}W$ and $t\bar{t}Z$ production at $\sqrt{s} = 13$ TeV are about 3–4 times higher than those at 8 TeV [1]. This, coupled with the higher integrated luminosity collected at 13 TeV collisions, allows for a much more accurate study of these processes. Precise measurements of the production cross section for $t\bar{t}W$ and $t\bar{t}Z$ are of particular interest because these topologies can receive sizeable contributions from new physics (NP) beyond the standard model (SM) [2, 3]. Furthermore, these processes form dominant backgrounds to several searches for NP, as well as to the measurements of SM processes, such as $t\bar{t}$ production in association with the Higgs boson ($t\bar{t}H$). In addition, $t\bar{t}Z$ production is the most sensitive process for directly measuring the coupling of the top quark to the Z boson. The Feynman diagrams for the dominant production mechanisms of these processes are shown in Fig. 1, to which the charge-conjugate states should be added.

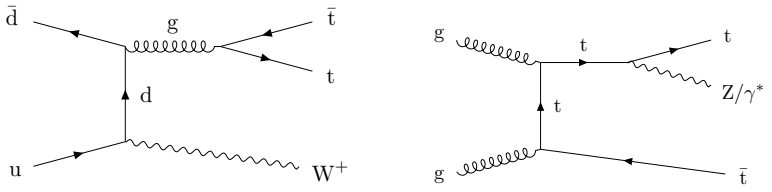


Figure 1: Representative leading-order Feynman diagrams for $t\bar{t}W$ and $t\bar{t}Z$ production at the LHC.

The $t\bar{t}Z$ cross section was measured by the CMS collaboration at $\sqrt{s} = 7$ TeV with a precision of $\approx 50\%$ [4]. At $\sqrt{s} = 8$ TeV CMS used multivariate techniques in events containing two, three, or four charged leptons to measure the $t\bar{t}W$ and $t\bar{t}Z$ cross sections with a precision of 30 and 25%, respectively [5, 6]. The $t\bar{t}Z$ process was observed with a significance of 6.4 standard deviations, and evidence for $t\bar{t}W$ production was found with a significance of 4.8 standard deviations. The ATLAS Collaboration analyzed events containing two and three charged leptons for its $t\bar{t}W$ measurement, and using two, three, and four charged leptons for the $t\bar{t}Z$ channel, achieving a similar precision [7]. In a more recent publication, the ATLAS Collaboration reported the first measurement of the $t\bar{t}W$ and $t\bar{t}Z$ production cross sections at $\sqrt{s} = 13$ TeV [8] with a significantly smaller data set than the one considered here.

In this paper we present measurements of the $t\bar{t}Z$ and $t\bar{t}W$ production cross sections at $\sqrt{s} = 13$ TeV with a data set corresponding to an integrated luminosity of 35.9 fb^{-1}. The measurements are performed using events in which at least one of the W bosons, originating from a top quark decay, further decays to a charged lepton and a neutrino, and the associated W or Z boson decays to a charged lepton and a neutrino or a charged lepton pair, where the charged lepton (ℓ) refers to an electron or a muon. The contribution from τ leptons are included through their decays to electrons and muons. The analysis is performed in three exclusive final states, in which events with two leptons of same charge, denoted as same-sign (SS) dileptons, are used to extract the $t\bar{t}W$ signal, while events with three or four charged leptons that include a lepton pair of opposite charge and same flavor (OSSF) are used to measure the $t\bar{t}Z$ signal yield. In addition to the individual $t\bar{t}W$ and $t\bar{t}Z$ cross section measurements, a fit is performed in all three final

states to simultaneously extract these cross sections. Furthermore, the results are interpreted in the context of an effective field theory to constrain the Wilson coefficients [9], which parameterize the strength of new physics interactions, for a set of selected dimension-six operators that might signal the presence of NP contributions in $\mathrm{t\bar{t}W}$ and $\mathrm{t\bar{t}Z}$ production.

2 The CMS detector

The central feature of the CMS apparatus is a superconducting solenoid of 6 m internal diameter, providing a magnetic field of 3.8 T. Within the solenoid volume are a silicon pixel and strip tracker, a lead tungstate crystal electromagnetic calorimeter (ECAL), and a brass and scintillator hadron calorimeter (HCAL), each composed of a barrel and two endcap sections. Forward calorimeters extend the pseudorapidity (η) coverage provided by the barrel and endcap detectors. Muons are detected in gas-ionization chambers embedded in the steel magnetic flux-return yoke outside the solenoid. A more detailed description of the CMS detector, together with a definition of the coordinate system used and the relevant kinematic variables, can be found in Ref. [10]. Events of interest are selected using a two-tiered trigger system [11]. The first level, composed of custom hardware processors, uses information from the calorimeters and muon detectors to select events, while the second level selects events by running a version of the full event reconstruction software optimized for fast processing on a farm of computer processors.

3 Event and object selection

Events are selected by online triggers that require the presence of at least one electron or muon, with transverse momentum, p_{T}, greater than 27 or 24 GeV, respectively. The selection efficiencies for the signal and background events that pass all requirements are found to be greater than 95 and 98% for the dilepton analysis and for the three- and four-lepton analyses, respectively.

The Monte Carlo (MC) simulations are used to estimate some of the backgrounds, as well as to calculate the selection efficiencies for the $\mathrm{t\bar{t}Z}$ and $\mathrm{t\bar{t}W}$ signal events. The simulated events for the $\mathrm{W}\gamma^*$, WW, tWZ, and for pairs of top quarks associated with a pair of bosons ($\mathrm{t\bar{t}VV}$, where V = W, Z, or H) processes, are performed at leading order (LO) in QCD, and for $\mathrm{t\bar{t}Z}$, $\mathrm{t\bar{t}W}$, tZq, tHq, tHW, WZ, WWZ, WZZ, ZZZ, $\mathrm{t\bar{t}}\gamma^*$, and $\mathrm{Z}\gamma^*$ final states at NLO in QCD using the MADGRAPH5_aMC@NLO v2.2.2 or v2.3.3 [12]. The NLO POWHEG v2 [13] generator is used for the production of the $\mathrm{t\bar{t}H}$ [14] and $\mathrm{q\bar{q}}\rightarrow$ ZZ [15, 16] processes, while the $\mathrm{gg} \rightarrow$ ZZ process is generated at LO in QCD with MCFM v7.0 [17]. The simulated samples of ZZ events are scaled to the cross sections calculated at next-to-next-to-leading order (NNLO) in QCD for $\mathrm{q\bar{q}} \rightarrow$ ZZ [18] (using a scaling K factor of 1.1) and for $\mathrm{gg} \rightarrow$ ZZ at NLO [19] (using $K = 1.7$). The NNPDF3.0LO [20] parton distribution functions (PDFs) are used for the simulation generated at LO and the NNPDF3.0NLO [20] PDF for those generated at NLO. Parton showering, hadronization, and the underlying event are simulated using PYTHIA v8.212 [21, 22] with the CUETP8M1 tune [23, 24]. The double counting of the partons generated with MADGRAPH5_aMC@NLO and those with PYTHIA is removed using the MLM [25] and the FxFx [26] matching schemes, in the LO and NLO generated events, respectively. All events are processed through a simulation of the CMS detector based on GEANT4 [27] and are reconstructed with the same algorithms as used for data. Simultaneous pp collisions in the same or nearby bunch crossings, referred to as pileup (PU), are also simulated. The PU distribution used in simulation is chosen to match the one observed in the data.

3

The theoretical cross sections for the $t\bar{t}W$ and $t\bar{t}Z$ signal processes are computed at NLO in QCD using MADGRAPH5_aMC@NLO and found to be 0.628 ± 0.082 and 0.839 ± 0.101 pb [1], respectively. These values are used to normalize the expected signal yields, as well as to rescale the measured signal strengths to obtain the final cross sections. In the calculation for $t\bar{t}Z$, the cross section corresponds to a phase space where the invariant mass of all pairs of leptons is required to be greater than 10 GeV.

Event reconstruction uses the CMS particle-flow (PF) algorithm [28] for particle reconstruction and identification. Because of PU, there can be far more than one collision vertex reconstructed per event. The reconstructed vertex for which the sum of the p_T of the physics objects is largest is chosen to be the primary pp interaction vertex. The physics objects here are the objects obtained by a jet finding algorithm [29, 30] applied to all charged tracks associated with this vertex, plus the missing transverse momentum (p_T^{miss}), which is computed as the magnitude of the vector sum of the p_T of all PF candidates.

Leptons are required to have $p_T > 10$ GeV and $|\eta| < 2.5\,(2.4)$ for electrons (muons) and must be isolated from the other particles produced in the collision. A relative isolation parameter, I_{rel}, is determined by a cone-based algorithm. For each electron (muon) candidate, a cone of $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} = 0.3\,(0.4)$ is constructed around the track direction at the event primary vertex, where $\Delta\eta$ and $\Delta\phi$ are the respective differences in pseudorapidity and azimuthal angle (in radians) relative to the lepton track. The scalar sum of the p_T of all PF particles within this cone is calculated, excluding the lepton candidate and any charged particle not originating from the selected primary vertex. Exclusion of such particles removes the PU contribution from the charged particles, and a correction is therefore still required for the neutral component. The average energy density deposited by neutral particles in the event, computed with the FASTJET [30, 31] program, is therefore subtracted from the neutral component to the sum of the p_T of particles in the cone. The quantity I_{rel} is then defined as the ratio of this corrected sum to the p_T of the lepton candidate. An electron candidate is selected if $I_{rel} < 0.1$ for all three analyses, while a muon candidate is selected if $I_{rel} < 0.25$ for the three- and four-lepton analyses, and if $I_{rel} < 0.15$ for the SS dilepton analysis. Consistency of the origination of the lepton from the primary vertex is enforced by requiring lepton transverse and longitudinal displacements from the primary vertex to be less than 0.05 and 0.1 cm, respectively. Additionally, the transverse impact parameter is required to be smaller than 4 standard deviations, where the impact parameter is the minimum spatial distance between the lepton trajectory and the primary vertex.

Jets are reconstructed by clustering PF candidates using the anti-k_T algorithm [29] with a distance parameter $R = 0.4$. The influence of PU is mitigated through a charged-hadron subtraction technique, which removes the energy of charged hadrons not originating from the primary vertex [32]. Jets are calibrated in simulation, and separately in data, accounting for energy deposits of neutral particles from PU and any nonlinear detector response. Calibrated jets with $p_T > 30$ GeV and $|\eta| < 2.4$ are selected for the analysis. Furthermore, jets formed with fewer than three PF candidates or with electromagnetic or hadronic energy fractions greater than 99% are vetoed. A selected jet can also overlap with selected leptons and lead thereby to some double counting. To prevent such cases, jets that are found within a cone of $\Delta R = 0.4$ around any of the signal leptons are removed from consideration.

A multivariate b tagging discriminator [33, 34] is used to identify jets that originate from the hadronization of b quarks (b jets). The selection criteria used in this analysis gives about 1% rate for tagging light-quark or gluon jets as b jets and a corresponding b tagging efficiency of around 70%, depending on the jet p_T and η.

4 Event selection

4.1 SS dilepton analysis

We measure the production rate of $\mathrm{t\bar{t}W}$ events in the decay channel that yields exactly two leptons with the same charge. Requiring the same electric charge for the two leptons retains only one third of the signal in the dilepton final state. However, this selection significantly improves the signal-to-background ratio, as SS lepton pairs are produced in SM processes with relatively small cross sections. The main backgrounds to this analysis originate from misreconstruction effects: misidentification of leptons from heavy-quark decays, hereafter called nonprompt leptons to distinguish them from prompt leptons originating from W and Z boson decays, and mismeasurement of the charge of one of the leptons in events with an oppositely charged lepton pair.

We select events with two SS leptons ($\mu\mu$, μe, ee), requiring the p_{T} of both leptons to be above 25 GeV. To avoid inefficiencies due to the trigger selection in the ee channel, the electron with higher p_{T} is required to have $p_{\mathrm{T}} > 40$ GeV. Events containing additional leptons passing looser identification and isolation requirements are vetoed. These loose identification and isolation criteria are the same as used to estimate the nonprompt background in data (see Section 5). The invariant mass of the two leptons must be greater than 12 GeV to suppress Drell–Yan (DY) and quarkonium processes. To suppress $\mathrm{Z} \to \mathrm{e^+e^-}$ events, the invariant mass of the two electrons is required to lie outside the 15 GeV window around the Z boson mass $M(\mathrm{Z})$ [35], followed by the requirement that $p_{\mathrm{T}}^{\mathrm{miss}} > 30$ GeV.

In order to distinguish these backgrounds from the signal, a multivariate analysis (MVA) has been developed. The MVA has been trained using the $\mathrm{t\bar{t}W}$ signal and the main background process, using events with at least two jets, one or more of which are identified as b jets. Among the observables examined as inputs to the MVA training, the following are found to provide the best discrimination between the signal and background: the number of jets, N_{j}, the number of b jets, N_{b}, the scalar sum of p_{T} of the jets, H_{T}, $p_{\mathrm{T}}^{\mathrm{miss}}$, the highest-$p_{\mathrm{T}}$ (leading) and the lowest-p_{T} (trailing) lepton p_{T}, the invariant mass calculated using $p_{\mathrm{T}}^{\mathrm{miss}}$ and p_{T} of each lepton, M_{T}, the leading and next-to-highest-p_{T} (subleading) jet p_{T}, and the separation ΔR between the trailing lepton and the nearest selected jet.

A boosted decision tree classifier with gradient boosting [36] is used as the MVA discriminant, and simulated events are split into equal training and testing samples. Figure 2 shows the kinematic distributions of variables used in the MVA, and Fig. 3 displays the output of the boosted decision tree classifier (D) for all background sources and the signal, scaled to the integrated luminosity of the analyzed data samples.

Events with $D > 0$ are selected to suppress the background from nonprompt leptons, and, for final signal extraction, they are split into two categories: $0 < D < 0.6$ and $D > 0.6$. These values are optimized to achieve the best expected sensitivity for $\mathrm{t\bar{t}W}$. Furthermore, the number of jets and b jets are also used to form five exclusive event categories that maximize signal significance. The categories are formed using events with $N_{\mathrm{j}} = 2$, 3, and >3. The latter two categories are further split according to the number of b jets, $N_{\mathrm{b}} = 1$ and $N_{\mathrm{b}} > 1$. Events with $D < 0$ are also used in the signal extraction procedure to constrain the uncertainties in the nonprompt lepton background.

Each of these categories is further split into two sets according to the total charge of the leptons: $\ell^+\ell^+$ or $\ell^-\ell^-$. This increases the sensitivity to the charge-asymmetric production of the signal ($\mathrm{t\bar{t}W^+}$ vs. $\mathrm{t\bar{t}W^-}$) resulting from the pp nature of the collision at the LHC, while the main backgrounds yield charge-symmetric dileptons. In total, we form 20 exclusive signal regions.

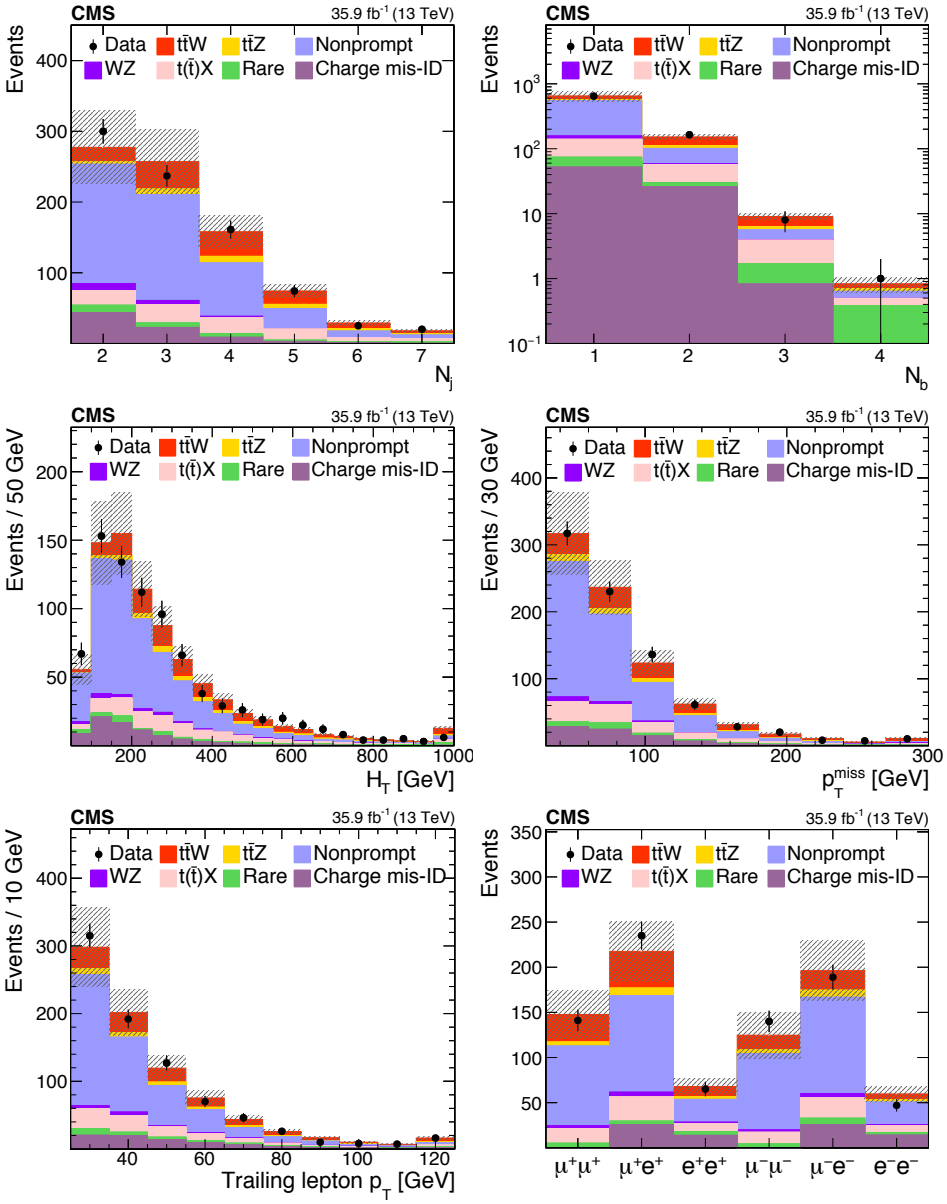


Figure 2: Distributions of different variables in data from the SS dilepton analysis, compared to the MC generated expectations. From left to right: jet and b jet multiplicity (upper), H_T and p_T^{miss} (center), trailing lepton p_T and event yields in each lepton-flavor combination (lower). The expected contributions from the different background processes are stacked, as well as the expected contribution from the signal. The shaded band represents the total uncertainty in the prediction of the background and the signal processes. See Section 5 for the definition of each background category.

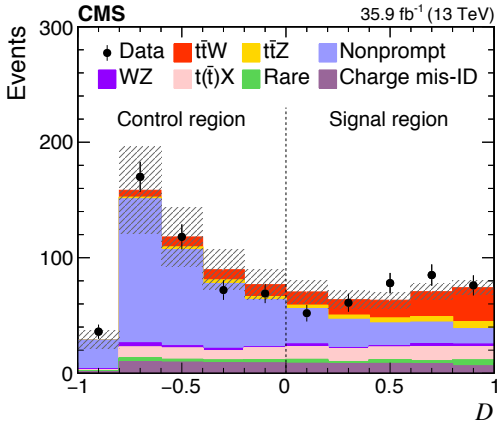


Figure 3: Distribution of the boosted decision tree classifier D for background and signal processes in the SS dilepton analysis. The expected contribution from the different background processes, and the signal as well as the observed data are shown. The shaded band represents the total uncertainty in the prediction of the background and the signal processes. See Section 5 for the definition of each background category.

4.2 Three-lepton analysis

The production rate of $\mathrm{t\bar{t}Z}$ events is measured in the final state with three leptons.

We select events that contain exactly three leptons ($\mu\mu\mu$, $\mu\mu$e, μee, or eee), requiring the leading, subleading, and trailing lepton p_{T} to be above 40, 20, and 10 GeV, respectively. To reduce backgrounds from multilepton processes that do not contain a Z boson, we require at least one OSSF lepton pair with invariant mass, $M(\ell\ell)$, consistent with the Z boson hypothesis, namely $|M(\ell\ell) - M(\mathrm{Z})| < 10\,\mathrm{GeV}$.

Signal events are expected to have at least four jets, two of which originate from b quarks. When the events pass the jet and b jet requirements defined in the previous section, one obtains a sample of events enriched in signal, with minimal background contribution. However, nearly 70% of the signal events fail the requirement of having four jets with two of them identified as b jets. We therefore make use of lower jet and b jet multiplicities to form nine exclusive event categories to include a larger fraction of the signal events. These nine categories are formed using events with $N_{\mathrm{j}} = 2$, 3, and > 3, where each jet multiplicity gets further split according to the b jet multiplicity, $N_{\mathrm{b}} = 0$, 1, and > 1.

Despite the larger background contamination, the $N_{\mathrm{j}} = 3$ categories, especially in bins with larger N_{b}, improve the signal sensitivity, as this category recovers signal efficiency for the jets that fall outside the acceptance. The $N_{\mathrm{j}} = 2$ category provides a background-dominated region that helps to constrain the background uncertainties. We use all nine signal regions to extract the signal significance and the cross section.

4.3 Four-lepton analysis

In addition to the three-lepton final state, events with four leptons are exclusively analyzed for the measurement of the $\mathrm{t\bar{t}Z}$ production rate.

The $\mathrm{t\bar{t}Z}$ events in this channel are characterized by the presence of two b jets, $p_{\mathrm{T}}^{\mathrm{miss}}$, and four leptons, two of which form an OSSF pair consistent with the Z boson mass. The event selection

is optimized to obtain high signal efficiency in simulation in order to profit from low expected background yields. Events with exactly four leptons that pass the lepton identification and isolation requirements described in Section 3 are selected. The leading lepton must have $p_T > 40\,GeV$ and the p_T of the remaining three leptons must exceed 10 GeV. The sum of the lepton charges must be zero, and the invariant mass of any lepton pair is required to be greater than 12 GeV. At least one OSSF lepton pair with an invariant mass $|M(\ell\ell) - M(Z)| < 20\,GeV$ must be present in the event. Events with $\mu\mu\mu\mu$, eeee, and $\mu\mu$ee final states, in which a second OSSF lepton pair consistent with the Z boson mass is found, are rejected. Events containing two jets are selected and split into two categories for signal extraction: one with zero b jets and the other with at least one b jet.

5 Background predictions

5.1 Background due to nonprompt leptons

Nonprompt leptons, i.e. leptons from heavy-flavor hadron decay, misidentified hadrons, muons from light meson decays, or electrons from unidentified photon conversions, are strongly rejected by the identification and the isolation criteria applied on electrons and muons. Nonetheless, a residual background from such leptons leaks into the analysis selection. Such backgrounds are mainly expected from $t\bar{t}$ production, in which one or two of the leptons originate from the leptonic W boson decays and an additional nonprompt lepton comes from the semileptonic decays of a b hadron, as well as from $Z \to \ell\ell$ events containing an additional misidentified lepton. These backgrounds are estimated using a data-based technique. From a control sample in data, we calculate the probability for a loosely identified nonprompt lepton to pass the full set of tight requirements, designated as the tight-to-loose ratio. For loose leptons we choose a relaxed isolation requirement, $I_{rel} < 1$, and additional electron identification requirements on the variables that distinguish prompt electrons from hadrons and photons which are misidentified as electrons. The tight-to-loose ratios are measured in a data control sample of QCD multijet events that are enriched in nonprompt leptons. This control sample consists of events with a single lepton and at least one jet, where the lepton and jets are separated by $\Delta R > 1$. We suppress the prompt lepton contamination, mostly from W+jets, by requiring $p_T^{miss} < 20\,GeV$ and $M_T < 20\,GeV$, where M_T is the transverse mass constructed using p_T^{miss} and the selected lepton. The residual prompt lepton contamination is subtracted using estimates from MC simulation. This subtraction is relevant only for the high-p_T leptons, and its effect on the total estimated background does not exceed a few percent. These tight-to-loose ratios are parametrized as functions of the η of the leptons and p_T^{cor}, with the latter calculated through corrections to lepton p_T as a function of the energy in the isolation cone. This definition has no impact on the p_T of the leptons that pass the isolation requirement, but modifies the p_T of those that fail, and extract thereby a more accurate value of true p_T [37]. The tight-to-loose ratios are then used together with the observed number of events in sideband regions. These sideband regions contain events that pass full event criteria in each analysis region, except that at least one of the leptons passes the loose selection but does not pass the tight selection. Each event in this region is assigned a weight as a function of the p_T and η of the loose lepton to account for the probability of the lepton to pass the tight selection.

We validate this technique using simulated events. The tight-to-loose ratios are first measured for electrons and muons in simulated multijet events, and applied in simulated $t\bar{t}$ and Z+jets events in the same way as in data, to extract predictions for the nonprompt background contribution. These predictions agree very well with the observed yields in simulation, not only for the integral yields, but also for distributions in all kinematic variables used to form the analysis

regions, including the boosted decision tree output D. Additionally, data control regions used in the signal-extraction regions and enriched in processes with nonprompt leptons, are formed to check any other potential sources of mismodeling. For the SS dilepton channel, we use the region with $D < 0$. Figure 4 shows the predicted background and observed data yields versus N_j and the p_T of the trailing lepton. Events in this region are also used in the signal extraction procedure for $t\bar{t}W$. The potential systematic effects for the extrapolation from $D < 0$ to $D > 0$ are studied in simulation and found to be negligible compared to other sources of uncertainty. For the three-lepton channel this control region is defined by either the absence of an OSSF lepton pair, or by the presence of an OSSF, with its invariant mass being at least 10 GeV away from $M(Z)$, and with at least one b jet present. This region is dominated by $t\bar{t}$ events in which both W bosons decay leptonically and an additional nonprompt lepton is present. Figure 5 shows the predicted and observed yields versus the flavor of the leptons, p_T^{miss}, N_j, and N_b. Both of these control regions show very good agreement between predicted and observed yields and for kinematic distributions that are relevant for the signal extraction.

Based on the extensive aforementioned validation in both data and simulated control samples, we conclude that a systematic uncertainty of 30% is appropriate for the prediction of the background from nonprompt leptons. The statistical uncertainties due to the limited number of observed events in the sideband regions of data are taken into account, and are often found to be larger or comparable to the systematic uncertainty.

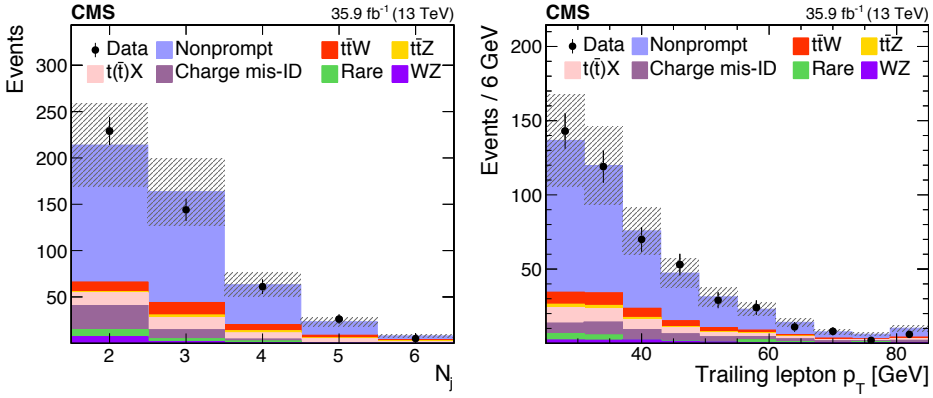


Figure 4: Distributions of the predicted and observed yields versus N_j (left) and p_T of the trailing lepton (right) in control regions enriched with nonprompt lepton background in the SS dilepton channel. The shaded band represents the total uncertainty in the prediction of the background and the signal processes. See Section 5 for the definition of each background category.

5.2 Background induced by the mismeasurement of the lepton charge

The charge mismeasurement rate for muons is negligible and background is significant only for the channels with at least one electron. This background is estimated with a partially data-based approach. The opposite-charge ee or eμ data events passing the full kinematic selection are weighted by the p_T- and η-dependent electron-charge misidentification probabilities. These probabilities are obtained from MC simulation. The charge mismeasurement rate in simulation is validated through a comparison with data. It is measured in DY events in MC and in data, where events are selected when the two SS electrons have an invariant mass that falls within a Z boson mass window, $76 < M(\ell\ell) < 106$ GeV. The measured electron charge misidentification

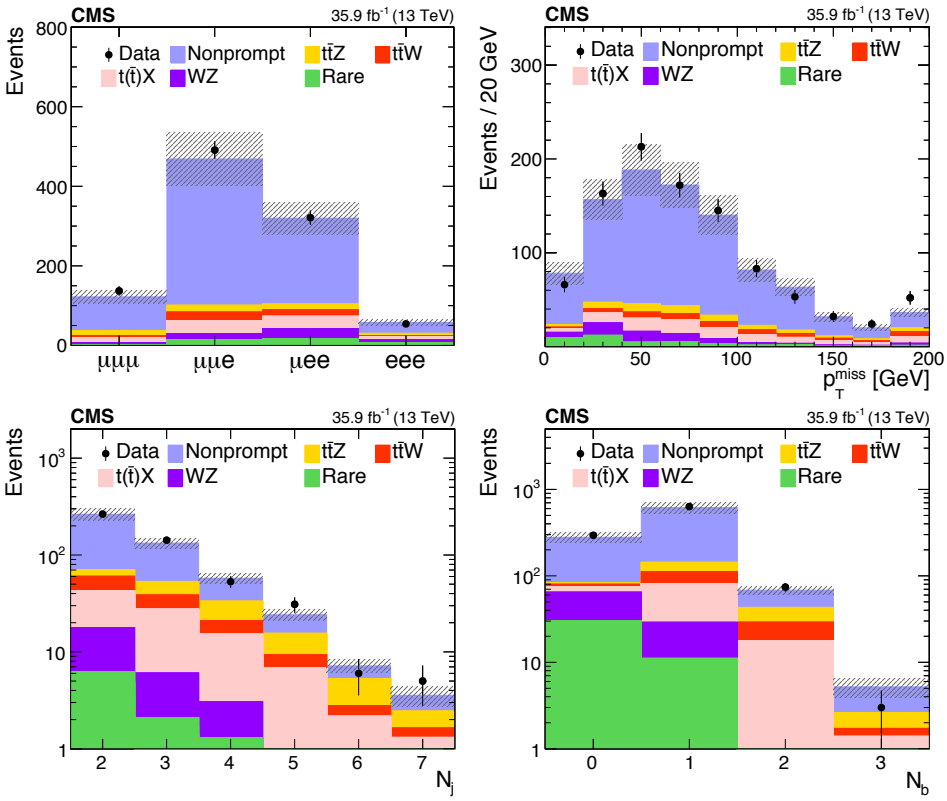


Figure 5: Distributions of the predicted and observed yields versus different three-lepton channels, $p_{\mathrm{T}}^{\mathrm{miss}}$ (upper panels), and jet and b jet multiplicity (lower panels) in control regions enriched with nonprompt lepton background. The shaded band represents the total uncertainty in the prediction of the background and the signal processes. See Section 5 for the definition of each background category.

rates in data and in DY simulation are in good agreement and vary from 4×10^{-5} for low-p_{T} electrons in the barrel region to 4×10^{-3} for high-p_{T} electrons in the endcap.

The process contributing to this category of background in signal regions is primarily $\mathrm{t\bar{t}}$ production. Based on the agreement in the charge mismeasurement rate between data and MC events, and the simulation studies of charge misidentification rate comparison between $\mathrm{t\bar{t}}$ and DY MC events, we assign a 20% systematic uncertainty in the estimation of this background [38].

5.3 Background due to WZ production

Kinematic distributions for the background from WZ events are taken from simulation. This background has the highest expected yields in the analysis region with no b-tagged jets. The data used for this analysis contain a substantial number of WZ events that can be isolated and compared with the MC predictions. We define a control region in a subset of the data with the following requirements: we select events with three leptons, with the same p_{T} thresholds as the ones used in the $\mathrm{t\bar{t}Z}$ selection, that have two leptons forming an OSSF pair with $|M(\ell\ell) - M(\mathrm{Z})| < 10\,\mathrm{GeV}$, less than two jets, and no b-tagged jets. Additionally $p_{\mathrm{T}}^{\mathrm{miss}}$ is required to be greater than 30 GeV, and M_{T}, constructed using this $p_{\mathrm{T}}^{\mathrm{miss}}$ and the lepton not used in the $M(\ell\ell)$

calculation, is required to be greater than 50 GeV.

This selection provides a data sample that is expected to be 85% pure in WZ events. Figure 6 shows the number of events as a function of M_{T}, lepton flavor, N_{j}, and $M(\ell\ell)$. The expected background from nonprompt leptons is measured from data using the method described above. The other background contributions are obtained from simulation. We observe overall agreement between data and the total expectation in all four-lepton channels and also in the kinematic distributions. The ratio of the total observed yield to the predicted one is found to be 0.94 ± 0.07, where the uncertainty reflects only statistical sources. With this level of agreement between the data and MC prediction, we proceed without applying any corrections to the WZ prediction obtained from the simulation. The statistical uncertainty in the ratio is propagated to the final prediction. We also study possible mismodeling of the WZ + heavy-flavor background at large b jet multiplicities. We find that the WZ contribution at high b jet multiplicities is mainly caused by the misidentification of light-flavor jets as b jets. The fraction of WZ events containing at least one b quark is predicted by the simulation to vary between 5 and 15% across all of the analysis categories. We apply scale factors to take into account the differences in b tagging efficiencies and misidentification probabilities between data and simulation [33, 34]. Once all the corrections are applied, we check the agreement between data and Z+jets simulated events as a function of N_{b} in OSSF dilepton events consistent with the $M(\mathrm{Z})$. Based on this study, we assign a 10% systematic uncertainty to the WZ background estimate, which covers the differences between data and expectations found in the control region. For the three-lepton analysis, an additional 20% uncertainty is introduced for regions with $N_{\mathrm{j}} > 3$. Other systematic uncertainties associated with the extrapolation from this control region to high N_{j} or N_{b} regions, such as jet energy scale and b tagging uncertainties, are considered separately.

5.4 Background due to $\mathrm{t}(\bar{\mathrm{t}})\mathrm{X}$ and other rare SM processes

The background events containing either multiple bosons or top quark(s) in association with a W, Z, or a Higgs boson are estimated from simulation scaled by their NLO cross section and normalized to the integrated luminosity. The backgrounds that have at least one top quark in final state, i.e. $\mathrm{t\bar{t}H}$, tWZ, tqZ, tHq, tHW, $\mathrm{t\bar{t}VV}$, and $\mathrm{t\bar{t}t\bar{t}}$, are denoted as $\mathrm{t}(\bar{\mathrm{t}})\mathrm{X}$, while all others, i.e. WW, ZZ, $\mathrm{W}\gamma^{*}$, $\mathrm{Z}\gamma^{*}$, WWW, WWZ, WZZ, ZZZ, are grouped into the rare SM processes category.

For background yields in the $\mathrm{t}(\bar{\mathrm{t}})\mathrm{X}$ category, we studied the theoretical and systematic uncertainties separately. The theoretical uncertainties for the inclusive cross section are around 10% [12, 39, 40]. Using the simulations, we study the effect of the changes made in renormalization and factorization scales (μ_{R} and μ_{F}), as well as the uncertainties from choice in PDF in the phase-space region relative to this analysis. From these studies we deduce an additional theoretical uncertainty of 2%. On the experimental side, to account for the differences in the lepton-selection efficiencies, b jet identification efficiencies, mistagging rate between the simulation and the data, we apply scale factors to the predictions obtained from simulations, and assign systematic uncertainties associated with these scale factors. These experimental uncertainties are estimated in each analysis category (see Section 6) and are applied in addition to the above-mentioned 10% uncertainty in the yield.

The rate for the backgrounds from rare SM processes, except ZZ, are assigned an overall 50% systematic uncertainty. This is motivated by the fact that these processes are not yet measured at the LHC and the uncertainties associated with the absence of higher-order effects might be large in the phase-space region relevant to this analysis. For the ZZ background, the consis-

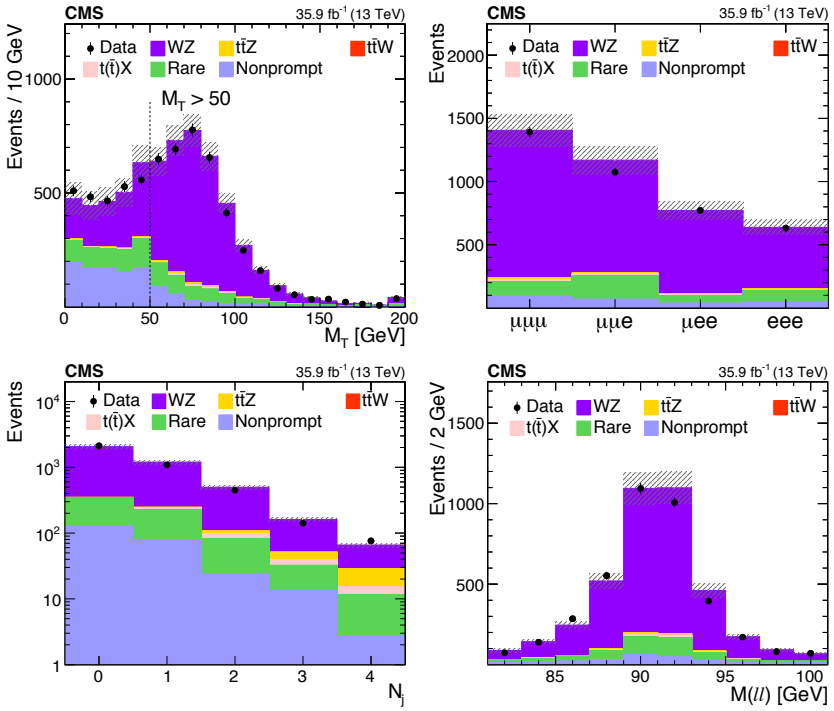


Figure 6: Distributions of the predicted and observed yields versus M_{T} (upper left), lepton flavor (upper right), jet multiplicity (lower left), and the reconstructed invariant mass of the Z boson candidates (lower right) in the WZ-enriched control region. The requirements on M_{T} and N_{j} are removed for the distributions of these variables. The shaded band represents the total uncertainty in the prediction of the background and the signal processes.

tency between data and simulation is validated in a ZZ-dominated background region. The events are selected following the first four steps mentioned in Section 4.2, in the given selection sequence, requiring two OSSF lepton pairs with an invariant mass within a 20 GeV window of $M(\mathrm{Z})$. The distributions of the expected and observed data yields in this ZZ enriched control region are shown in Fig. 7. The ZZ control region, which is better than 95% pure in ZZ events, shows good agreement between data and simulation in events with extra jets. Based on this study in the four-lepton control region, as well as considering the studies done for the WZ background at high jet multiplicities, we assign a 20% systematic uncertainty. Additional experimental uncertainties, as previously described for the $\mathrm{t}(\bar{\mathrm{t}})\mathrm{X}$ and WZ backgrounds, are also applied to the ZZ background.

6 Signal acceptance and systematic uncertainties

The uncertainty in the integrated luminosity is 2.5% [41]. Simulated events are reweighted according to the distribution of the true number of interactions at each bunch crossing. The uncertainty in the total inelastic pp cross section, which affects the PU estimate, is 5% [42] and it leads to a 1–2% uncertainty in the expected yields.

We measure the trigger efficiencies in a data sample independent from the one used for the signal selection, as well as in simulation. These efficiencies are measured for each channel sep-

12

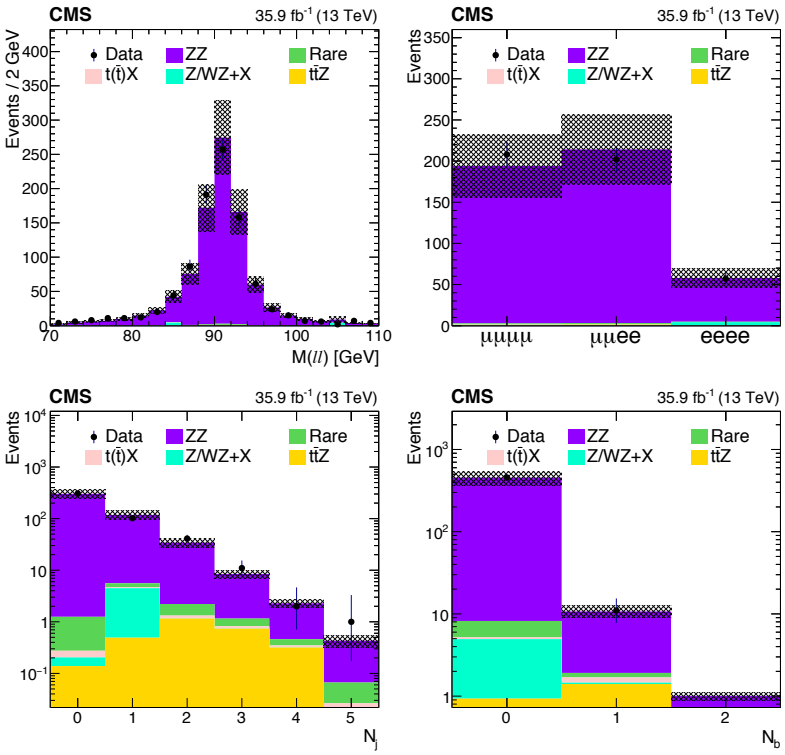


Figure 7: Comparison of data with MC predictions for the mass of the Z boson candidate (upper left), event yields (upper right), jet multiplicity (lower left) and b jet multiplicity (lower right) in a ZZ-dominated background control region. The shaded band represents the total uncertainty in the prediction of the background and the signal processes.

arately and parametrized as a function of lepton p_{T} and η. The overall efficiency for the SS dilepton channel is higher than 95% and that for the three- and four-lepton analyses is greater than 98%. The trigger efficiencies measured in simulation agree within 1% with the measurements in data, with an exception of the SS dimuon channel, in which the difference reaches 3%. The event yields in simulation are therefore scaled to match the trigger efficiencies in data. The systematic uncertainty due to this scaling is 2–4% depending on the channel.

Reconstructed lepton selection efficiencies are measured using a "tag-and-probe" method [43, 44] in bins of lepton p_{T} and η, and are higher than 65 (96)% for electrons (muons). These measurements are performed separately in data and in simulation. The differences between these two measurements are typically around 1–4% per lepton, which corresponds to 3–7% for all leptons in the event. The systematic uncertainties related to this source vary between 2 and 7%.

Uncertainties in the jet energy calibrations are estimated by shifting the energy of jets in the simulation up and down by one standard deviation. Depending on p_{T} and η, the uncertainty in jet energy scale changes by 2–5% [45, 46]. For the signal and backgrounds modelled through simulation, the resulting uncertainty is determined by the observed differences in yields with and without the shift in jet energies. The same technique is used to calculate the uncertainties caused by the jet energy resolution, for which the uncertainty is found to be 1–6%. These uncertainties are also propagated in the $p_{\mathrm{T}}^{\mathrm{miss}}$ variable, and the resulting uncertainty in signal

selection is found to be around 1%. The b tagging efficiency in the simulation is corrected using scale factors determined from data [33, 34]. These contribute with an uncertainty of about 2–5% on the predicted yields, which depend on p_T, η and jet and b-tag multiplicity.

To estimate the theoretical uncertainties due to μ_R and μ_F choices, each of these parameters is varied independently up and down by a factor of 2, ignoring the anti-correlated variations. For the acceptance uncertainties, the envelope of the results is used as an uncertainty in each search bin, and found not to exceed 2%. The different replicas in the NNPDF30 PDF set [20] are used to estimate the corresponding uncertainty in acceptance, which is typically less than 1%.

The theoretical uncertainty in the cross sections for top quark (pair) production in association with a Higgs boson or a vector boson is 11% [1]. For the WZ and ZZ backgrounds, the overall uncertainty in the cross section is 10%, with additional uncertainties at large jet multiplicities. Rare SM processes are assigned a 50% systematic uncertainty. All of the experimental uncertainties described above are evaluated for each process in all analysis categories. A 20% systematic uncertainty is assigned to the charge-misidentified background. The uncertainty in the nonprompt lepton contribution in the SS dilepton and three-lepton analyses is 30%, for which the statistical uncertainty in the observed yields in the sideband region is also taken into account.

The theoretical uncertainties for individual backgrounds as well as the systematic uncertainties for the nonprompt background are uncorrelated, but correlated across the analysis categories. The different sources of experimental uncertainty are correlated across the analysis categories and among the background and signal predictions. The statistical uncertainties from the limited number of events in MC simulation and from the data events in the sideband regions are considered fully uncorrelated.

The impact of different sources of systematic uncertainty is estimated by fixing the nuisance parameter corresponding to each uncertainty one at a time and evaluating the decrease in the total systematic uncertainty. Uncertainties associated with the integrated luminosity, lepton identification, trigger selection efficiencies, nonprompt lepton, and $t(\bar{t})X$ backgrounds have the greatest effect on both the $t\bar{t}W$ and the $t\bar{t}Z$ cross section measurements. The full set of systematic uncertainties is shown in Table 1.

7 Results

As described in Section 4, the data are analyzed in three exclusive channels according to the number of leptons in the final state: SS dilepton, three- and four-lepton events. Each channel is further categorized according to the number of jets and b-tagged jets. The predicted SM background and signal yields, and the observed data are shown in Figs. 8 and 9, and in Tables 2–5, for each of the above categories, respectively. In general, we find good agreement between the predicted yields and the observed data, except for some excess of events accumulated in the $N_j = 2, 3$ and $N_b > 1$ category of the three-lepton channel. Extensive studies were performed to ensure the robustness of the estimated background yields in this region. No hints of a missing or underestimated background were found; therefore, we attribute this excess to a statistical fluctuation in data. In Figs. 10 and 11, various kinematic distributions in the predicted and observed yields are presented in $t\bar{t}W$ and $t\bar{t}Z$ signal-enriched regions: SS dileptons with $N_j > 2$ and $N_b > 1$, and three-lepton events with $N_j > 2$ and $N_b > 0$, respectively.

The statistical procedure to extract the cross section is detailed in Refs. [47–50]. The observed yields and background estimates in each analysis category, described in Section 4, and the sys-

14

Table 1: Summary of the sources of uncertainties, their magnitudes, and their effects in the final measurement. The first column indicates the source of the uncertainties, while the second column shows the corresponding input uncertainty on each background source and the signal. The third and fourth columns show the resulting uncertainties in the respective $t\bar{t}W$ and $t\bar{t}Z$ cross sections.

Source	Uncertainty from each source (%)	Impact on the measured $t\bar{t}W$ cross section (%)	Impact on the measured $t\bar{t}Z$ cross section (%)
Integrated luminosity	2.5	4	3
Jet energy scale and resolution	2–5	3	3
Trigger	2–4	4–5	5
B tagging	1–5	2–5	4–5
PU modeling	1	1	1
Lepton ID efficiency	2–7	3	6–7
Choice in μ_R and μ_F	1	<1	1
PDF	1	<1	1
Nonprompt background	30	4	<2
WZ cross section	10–20	<1	2
ZZ cross section	20	—	1
Charge misidentification	20	3	—
Rare SM background	50	2	2
$t(\bar{t})X$ background	10–15	4	3
Stat. unc. in nonprompt background	5–50	4	2
Stat. unc. in rare SM backgrounds	20–100	1	<1
Total systematic uncertainty	—	14	12

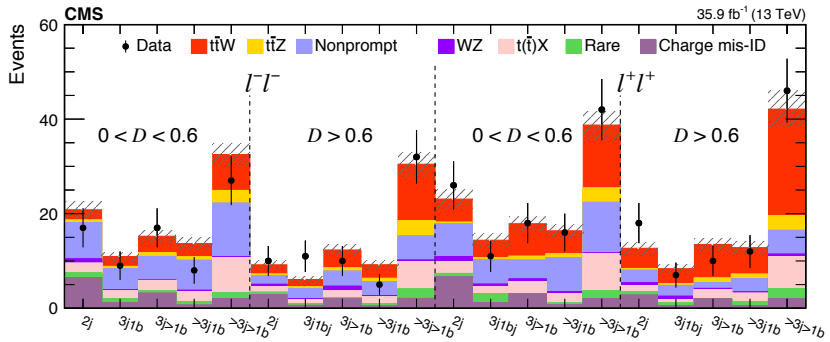


Figure 8: Predicted signal and background yields, as obtained from the fit, compared to observed data in the SS dilepton analysis. The hatched band shows the total uncertainty associated with the signal and background predictions, as obtained from the fit.

tematic uncertainties described in Section 6 are used to construct a binned likelihood function $L(r,\theta)$ as a product of Poisson probabilities of all bins. The parameter r is the signal-strength modifier and θ represents the full suite of nuisance parameters. The signal strength parameter $r = 1$ corresponds to a signal cross section equal to the SM prediction, while $r = 0$ corresponds to the background-only hypothesis.

The test statistic is the profile likelihood ratio, $q(r) = -2L(r,\hat{\theta}_r)/L(\hat{r},\hat{\theta})$, and asymptotic approximation is used to extract the fitted cross section, the associated uncertainties, and the significance of the observation of the signal process [47–50], where $\hat{\theta}_r$ reflects the values of the nuisance parameters that maximize the likelihood function for signal strength r. The quantities $\hat{r}$ and $\hat{\theta}$ are the values that simultaneously maximize L.

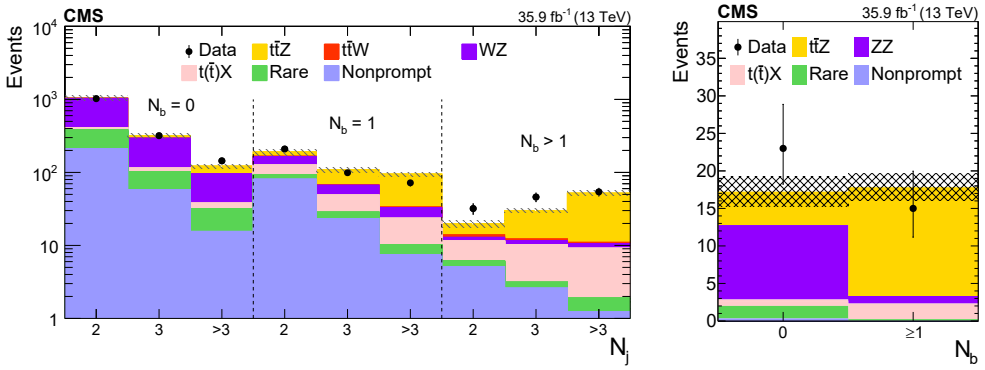


Figure 9: Predicted signal and background yields, as obtained from the fit, compared to observed data in $N_j = 2$, 3, and > 3 categories in the three-lepton analysis (left), and in $N_b = 0$, 1 categories in the four-lepton analysis (right). The hatched band shows the total uncertainty associated with the signal and background predictions, as obtained from the fit.

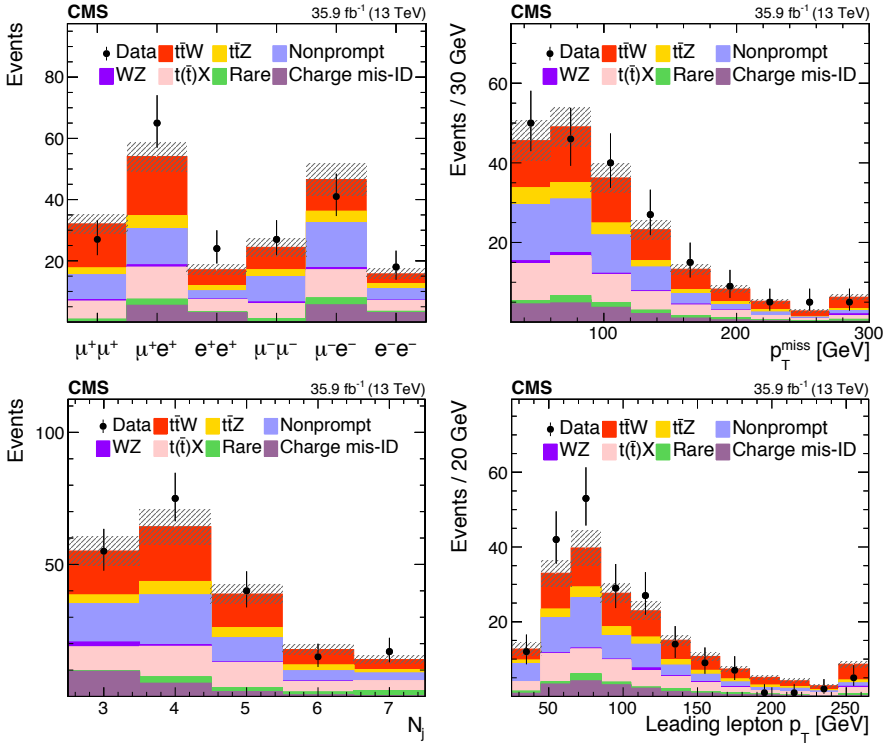


Figure 10: Predicted signal and background yields, as obtained from the fit, compared to observed data versus the flavor and the charge combination of leptons (upper left), $p_{\mathrm{T}}^{\mathrm{miss}}$ (upper right), jet multiplicity (lower left), and the p_{T} of the leading lepton (lower right) in the SS dilepton channel with at least three jets and at least two b jets. The last bin in each distribution includes the overflow events, and the hatched band shows the total uncertainty associated with the signal and background predictions, as obtained from the fit.

16

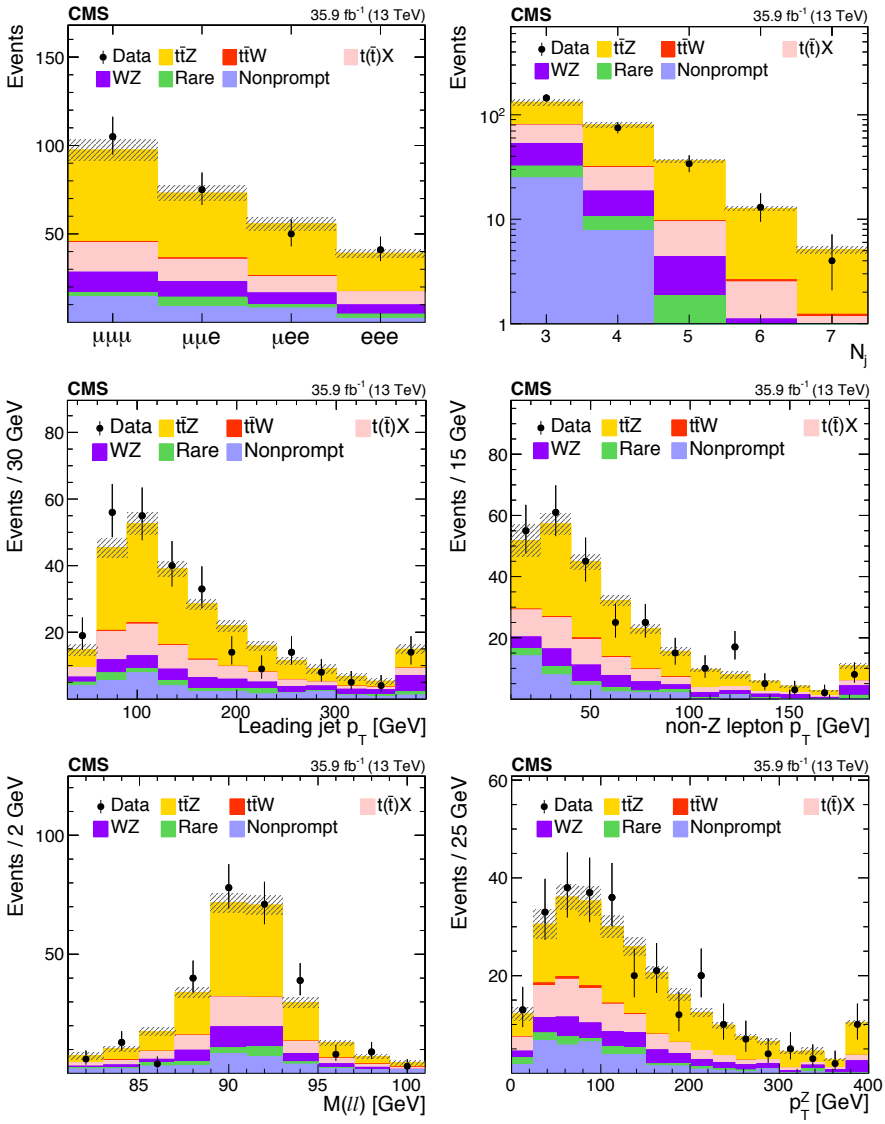


Figure 11: Predicted signal and background yields, as obtained from the fit, compared to observed data in the three-lepton channel for events containing at least three jets and at least one b jet. From left to right: the lepton flavor and jet multiplicity (upper), p_{T} of the leading jet and the lepton not used to form Z (central), and invariant mass of the OSSF lepton pair and p_{T} of the reconstructed Z boson (lower). The last bin in each distribution includes the overflow events, and the hatched band shows the total uncertainty associated with the signal and background predictions, as obtained from the fit.

The measurement of the individual cross sections for $\mathrm{t\bar{t}W}$ and $\mathrm{t\bar{t}Z}$ is performed using the events in the SS dilepton, and the three- and four-lepton categories, respectively, while the $\mathrm{t\bar{t}W^{+}}$($\mathrm{t\bar{t}W^{-}}$) signal extraction is performed using the SS dilepton category with $\ell^{+}\ell^{+}$($\ell^{-}\ell^{-}$). The summary of the expected and observed signal significances for each of these processes is given in Table 6. We find an expected (observed) signal significance of 4.5 (5.3) standard deviations in the SS

Table 2: Predicted signal and background yields, as obtained from the fit, compared to observed data in the SS dilepton channel for the $D < 0$ region, i.e. the nonprompt lepton control region. The total uncertainty obtained from the fit is also shown.

Process	$N_j = 2$	$N_j = 3$	$N_j > 3$
Nonprompt	136.5 ± 13.9	110.3 ± 11.3	57.3 ± 6.1
Total background	192.1 ± 15.6	137.7 ± 11.7	74.0 ± 6.4
$t\bar{t}W$	13.1 ± 0.3	17.6 ± 0.3	13.8 ± 0.3
$t\bar{t}Z$	1.6 ± 0.4	3.1 ± 0.7	4.4 ± 1.0
Total	206.8 ± 15.7	158.4 ± 11.8	92.3 ± 6.5
Observed	229	144	92

Table 3: Predicted signal and background yields, as obtained from the fit, compared to observed data in the SS dilepton final state. The total uncertainty obtained from the fit is also shown.

		N_j	N_b	Background	$t\bar{t}W$	$t\bar{t}Z$	Total	Observed
$\ell^-\ell^-$	$0 < D < 0.6$	2	>0	18.1 ± 1.8	2.2 ± 0.4	0.5 ± 0.1	20.8 ± 1.9	17
		3	1	8.3 ± 0.9	2.1 ± 0.4	0.5 ± 0.1	10.9 ± 0.9	9
			>1	10.9 ± 1.1	3.5 ± 0.6	0.8 ± 0.1	15.2 ± 1.3	17
		>3	1	10.1 ± 1.1	2.8 ± 0.5	0.7 ± 0.2	13.7 ± 1.3	8
			>1	22.2 ± 2.0	7.6 ± 1.2	2.7 ± 0.4	32.5 ± 2.4	27
	$D > 0.6$	2	>0	6.8 ± 0.9	2.0 ± 0.3	0.4 ± 0.1	9.2 ± 0.9	10
		3	1	4.1 ± 0.6	1.6 ± 0.3	0.3 ± 0.1	6.1 ± 0.6	11
			>1	7.8 ± 0.9	3.8 ± 0.6	0.7 ± 0.1	12.3 ± 1.1	10
		>3	1	5.6 ± 0.7	2.9 ± 0.5	0.7 ± 0.2	9.2 ± 0.9	5
			>1	15.3 ± 1.5	12.0 ± 1.9	3.2 ± 0.5	30.5 ± 2.5	32
$\ell^+\ell^+$	$0 < D < 0.6$	2	>0	17.9 ± 1.8	4.9 ± 0.8	0.3 ± 0.1	23.1 ± 2.0	26
		3	1	10.2 ± 1.3	3.7 ± 0.6	0.4 ± 0.1	14.4 ± 1.4	11
			>1	10.2 ± 1.2	6.9 ± 1.1	0.8 ± 0.2	17.9 ± 1.6	18
		>3	1	10.7 ± 1.2	4.9 ± 0.8	0.8 ± 0.2	16.4 ± 1.4	16
			>1	22.4 ± 2.0	13.3 ± 2.2	3.0 ± 0.5	38.7 ± 3.0	42
	$D > 0.6$	2	>0	8.0 ± 1.1	4.3 ± 0.7	0.4 ± 0.1	12.7 ± 1.3	18
		3	1	4.8 ± 0.7	3.2 ± 0.5	0.3 ± 0.1	8.4 ± 0.9	7
			>1	5.4 ± 0.7	7.1 ± 1.2	1.0 ± 0.2	13.5 ± 1.4	10
		>3	1	6.3 ± 0.8	5.6 ± 0.9	0.9 ± 0.2	12.8 ± 1.2	12
			>1	16.5 ± 1.5	22.5 ± 3.7	3.1 ± 0.5	42.1 ± 4.0	46

dilepton channel, and 4.7 (4.5) standard deviations in the four-lepton channel, while in three-lepton channel both the expected and the observed significances are found to be much larger than 5 standard deviations. The expected (observed) signal significances for $t\bar{t}W^+$ and $t\bar{t}W^-$ processes are calculated as well, being 4.2 (5.5) and 2.4 (2.3), respectively.

The measured signal strength parameters are found to be $1.23^{+0.19}_{-0.18}\,(\text{stat})^{+0.20}_{-0.18}\,(\text{syst})^{+0.13}_{-0.12}\,(\text{theo})$ for $t\bar{t}W$, and $1.17\,^{+0.11}_{-0.10}\,(\text{stat})^{+0.14}_{-0.12}\,(\text{syst})^{+0.11}_{-0.12}\,(\text{theo})$ for $t\bar{t}Z$. These parameters are used to multiply the corresponding theoretical cross sections for $t\bar{t}W$ and $t\bar{t}Z$ mentioned in Section 3, to obtain the measured cross sections for $t\bar{t}W$ and $t\bar{t}Z$:

$$\sigma(pp \to t\bar{t}W) = 0.77^{+0.12}_{-0.11}\,(\text{stat})^{+0.13}_{-0.12}\,(\text{syst})\,\text{pb},$$

Table 4: Predicted signal and background yields, as obtained from the fit, compared to observed data in the three-lepton final state. The total uncertainty obtained from the fit is also shown.

N_b	N_j	Background	$t\bar{t}W$	$t\bar{t}Z$	Total	Observed
	2	1032.8 ± 77.1	0.9 ± 0.1	18.2 ± 3.2	1051.9 ± 77.2	1022
0	3	293.5 ± 21.4	0.4 ± 0.1	22.3 ± 3.9	316.3 ± 21.8	318
	>3	95.4 ± 7.4	0.3 ± 0.1	26.1 ± 4.6	121.8 ± 8.7	144
	2	164.6 ± 17.8	1.9 ± 0.3	24.3 ± 4.3	190.7 ± 18.3	209
1	3	66.6 ± 6.7	0.9 ± 0.2	41.2 ± 7.2	108.7 ± 9.8	99
	>3	32.8 ± 3.3	0.8 ± 0.1	61.3 ± 10.8	94.9 ± 11.3	72
	2	12.9 ± 2.4	1.0 ± 0.2	5.9 ± 1.0	19.8 ± 2.6	32
>1	3	11.6 ± 1.7	0.6 ± 0.1	17.9 ± 3.2	30.1 ± 3.6	46
	>3	10.6 ± 1.6	0.4 ± 0.1	41.0 ± 7.2	52.0 ± 7.4	54

Table 5: Predicted signal and background yields, as obtained from the fit, compared to observed data in the four-lepton final state. The total uncertainty obtained from the fit is also shown.

Process	$N_b = 0$	$N_b > 0$
Total background	12.8 ± 2.0	3.3 ± 0.3
$t\bar{t}Z$	4.5 ± 0.6	14.5 ± 1.8
Total	17.2 ± 2.0	17.8 ± 1.8
Observed	23	15

Table 6: Summary of expected and observed significances (in standard deviations) for $t\bar{t}W$ and $t\bar{t}Z$.

Channel	Expected significance	Observed significance
SS dilepton $\ell^-\ell^-$ ($t\bar{t}W^-$)	2.4	2.3
SS dilepton $\ell^+\ell^+$($t\bar{t}W^+$)	4.2	5.5
SS dilepton $\ell^\pm\ell^\pm$ ($t\bar{t}W^\pm$)	4.5	5.3
Three-lepton ($t\bar{t}Z$)	>5.0	>5.0
Four-lepton ($t\bar{t}Z$)	4.7	4.5
Three- and four-lepton combined ($t\bar{t}Z$)	>5.0	>5.0

$$\sigma(pp \to t\bar{t}Z) = 0.99^{+0.09}_{-0.08}\,(\text{stat})^{+0.12}_{-0.10}\,(\text{syst})\,\text{pb}.$$

The measured cross sections for the $t\bar{t}W^+$ and $t\bar{t}W^-$ processes are:

$$\sigma(pp \to t\bar{t}W^+) = 0.58 \pm 0.09\,(\text{stat})^{+0.09}_{-0.08}\,(\text{syst})\,\text{pb},$$

$$\sigma(pp \to t\bar{t}W^-) = 0.19 \pm 0.07\,(\text{stat}) \pm 0.06\,(\text{syst})\,\text{pb}.$$

The individual measured cross sections for $t\bar{t}W$ and $t\bar{t}Z$, as well as the results of a simultaneous fit for these cross sections in all three analysis categories, SS dilepton, three-lepton, and four-lepton, are summarized in Fig. 12. The corresponding 68 and 95% confidence level (CL) contours and intervals are also shown. The cross section extracted for $t\bar{t}Z$ from the simultaneous fit is identical to the one obtained from the individual measurement, while for $t\bar{t}W$ the simultaneous fit result is shifted down by about 6%, which is smaller than the total systematic uncertainty. This is because the fitted value for the nonprompt background contribution in the three-lepton channel is 9% higher than the nominal value, so the fitted nonprompt yields in the SS dilepton channel are higher in the combined fit compared to the one in the individual fit.

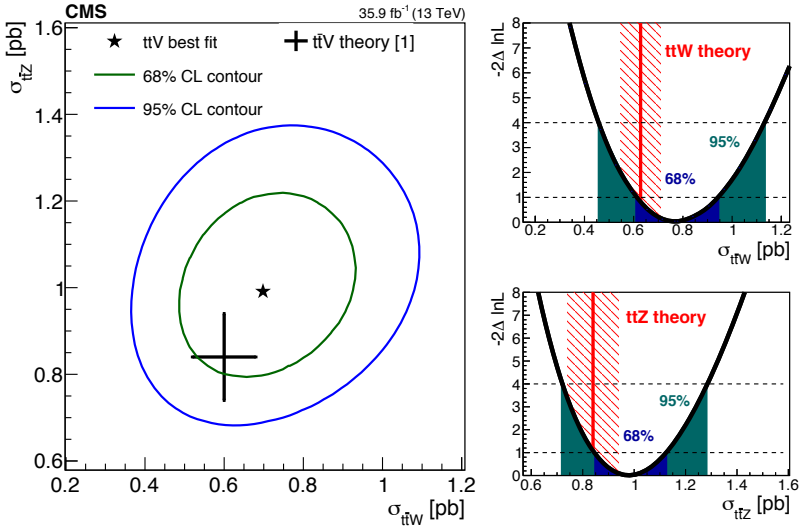


Figure 12: Result of the simultaneous fit for $t\bar{t}W$ and $t\bar{t}Z$ cross sections (denoted as star), along with its 68 and 95% CL contours are shown on the left panel. The right panel presents the individual measured cross sections along with the 68 and 95% CL intervals and the theory prediction [1] with their respective uncertainties for $t\bar{t}W$ and $t\bar{t}Z$.

8 Effective field theory interpretation

Within the framework of effective field theory, cross section measurements can be used to search for NP in a model-independent way at energy scales that are not yet experimentally accessible. Using this approach, the SM Lagrangian is extended with higher-order operators that correspond to combinations of SM fields. The extended Lagrangian is a series expansion in the inverse of the energy scale of the NP, $1/\Lambda$ [51], hence operators are suppressed as long as Λ is large compared with the experimentally-accessible energy.

The effective Lagrangian is (ignoring the single dimension-five operator, which violates lepton number conservation [51])

$$\mathcal{L}_{\text{eff}} = \mathcal{L}_{\text{SM}} + \frac{1}{\Lambda^2}\sum_i c_i \mathcal{O}_i + \cdots, \tag{1}$$

where $\mathcal{L}_{\text{SM}}$ is the dimension-four SM Lagrangian, $\mathcal{O}_i$ are dimension-six operators, and the ellipsis symbol represents higher-dimension operators. The dimensionless Wilson coefficients c_i parameterize the strength of the NP interaction.

Assuming baryon and lepton number conservation, there are fifty-nine independent dimension-six operators [52]. Thirty-nine of these operators were chosen for study in Ref. [53] because they include at least one Higgs field; the four-fermion operators were omitted. Constraints on the Wilson coefficients of some dimension-six operators have been reported in Refs. [2, 6, 54–60].

To investigate the effects of NP on any given process, it is necessary to calculate the expected cross section as a function of the Wilson coefficients. The matrix element can be written as the sum of SM and NP components:

$$\mathcal{M} = \mathcal{M}_0 + \sum_i c_i \mathcal{M}_i. \tag{2}$$

In this work, we consider one operator at a time. The cross section is proportional to the square

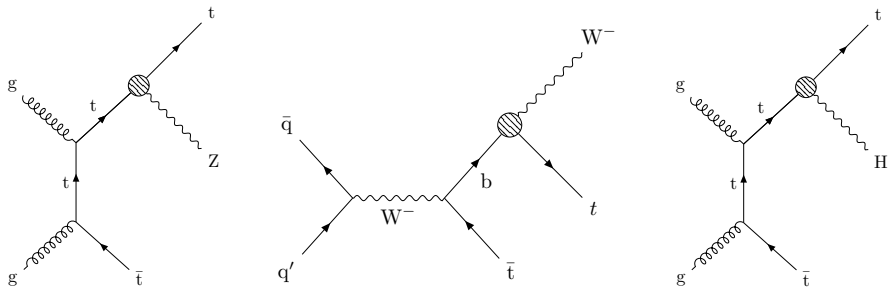


Figure 13: Feynman diagrams representing some of the most significant NP contributions to the $t\bar{t}Z$, $t\bar{t}W$, and $t\bar{t}H$ processes.

of the matrix element, and has the following structure [57]:

$$\begin{aligned}\sigma_{\text{SM+NP}}(c_i) &\propto |\mathcal{M}|^2 \\ &\propto s_0 + s_{1i}c_i + s_{2i}c_i^2.\end{aligned} \tag{3}$$

The coupling structures s_0, s_{1i}, and s_{2i} are constants which can be determined by evaluating the cross section for at least three values of c_i. Note that while $\sigma(c_i)$ is always quadratic, the minimum is not constrained to appear at the SM value ($c_i = 0$), and in cases of destructive interference with the SM, it is possible to have $\sigma_{\text{SM+NP}}(c_i) < \sigma_{\text{SM}}$.

NP effects on $t\bar{t}W$ and $t\bar{t}Z$ are considered. Because $t\bar{t}H$ is sizeable background to $t\bar{t}W$, and the NP effects on $t\bar{t}H$ are considered as well, as they cannot be disentangled from NP effects on $t\bar{t}W$. The range of Wilson coefficient values to study is chosen such that $|c_i| < (4\pi)^2$ [61]. The dimension-six operators are encoded using the the FeynRules [62] implementation from Ref. [53], and we follow their notation and operator-naming scheme throughout this work. This implementation assumes flavor-independent fermion couplings. Because the W and Z boson couplings to light quarks are highly constrained by other measurements, i.e. inclusive W or Z cross section measurements, we removed all NP couplings to the first two generations. This modified implementation is used in MADGRAPH5_aMC@NLO [63] to evaluate the cross section $\sigma_{\text{SM+NP}}$ expected due to both SM and NP effects at LO, with no constraints on the number of allowed QCD or electroweak vertices, for 30 values of c_i, with all other couplings set to their SM values. We then fit those points with a quadratic function (see Eq. (3)) to determine $\sigma_{\text{SM+NP}}(c_i)$.

The signal strength $r_{t\bar{t}Z}(c_i)$ is defined as the ratio of $\sigma_{\text{SM+NP, }t\bar{t}Z}(c_i)$ to $\sigma_{\text{SM+NP, }t\bar{t}Z}(0)$, and similarly for $t\bar{t}W$ and $t\bar{t}H$. We use this to construct a profile likelihood test statistic $q(c_i)$. The likelihood statistic is maximized to find the asymptotic best fit c_i, similarly to the procedure described in Section 7. Each coupling is profiled with the other couplings set to their SM values.

Of the thirty-nine operators in Ref. [53], we choose not to consider operators that do not affect $t\bar{t}W$, $t\bar{t}Z$, or $t\bar{t}H$. The expected 95% CL interval is calculated for the remaining 24 operators. We also exclude from consideration operators that produce large effects in better-measured processes for Wilson coefficient values to which our measurement is sensitive. To accomplish this, we require that the cross section for each of $t\bar{t}$, WW, ZZ, WZ, and inclusive Higgs boson production is not modified by more than 70% within our expected 95% CL interval. Finally, we do not include any operators that produce a significant effect on background yields (as described in Section 5) other than $t\bar{t}H$, as these can be studied more effectively in other signal regions.

Table 7: Expected 68% and 95% CL intervals for selected Wilson coefficients.

Wilson coefficient	68% CL [TeV^{-2}]	95% CL [TeV^{-2}]
$\bar{c}_{uW}/\Lambda^2$	$[-1.6, 1.5]$	$[-2.2, 2.2]$
$\lvert\bar{c}_H/\Lambda^2 - 16.8\,\text{TeV}^{-2}\rvert$	$[3.7, 23.4]$	$[0, 28.7]$
$\widetilde{c}_{3G}/\Lambda^2$	$[-0.5, 0.5]$	$[-0.7, 0.7]$
$\bar{c}_{3G}/\Lambda^2$	$[-0.3, 0.7]$	$[-0.5, 0.9]$
$\bar{c}_{uG}/\Lambda^2$	$[-0.9, -0.8]$ and $[-0.3, 0.2]$	$[-1.1, 0.3]$
$\lvert\bar{c}_{uB}/\Lambda^2\rvert$	$[0, 1.5]$	$[0, 2.1]$
$\bar{c}_{Hu}/\Lambda^2$	$[-9.2, -6.5]$ and $[-1.6, 1.1]$	$[-10.1, 2.0]$
$\bar{c}_{2G}/\Lambda^2$	$[-0.7, 0.4]$	$[-0.9, 0.6]$

Eight operators satisfy the above requirements, and constraints on their Wilson coefficients, $\bar{c}_{uW}$, $\bar{c}_H$, $\widetilde{c}_{3G}$, $\bar{c}_{3G}$, $\bar{c}_{uG}$, $\bar{c}_{uB}$, $\bar{c}_{Hu}$, and $\bar{c}_{2G}$ are reported here. Feynman diagrams representing some of the most significant NP contributions to the $t\bar{t}Z$, $t\bar{t}W$, and $t\bar{t}H$ processes are shown in Fig. 13.

The expected CL intervals for the selected Wilson coefficients are summarized in Table 7. Observed best fit values and CL intervals are summarized in Table 8. For three representative operators, the calculated signal strengths $r_{t\bar{t}Z}(c_i)$, $r_{t\bar{t}W}(c_i)$, and $r_{t\bar{t}H}(c_i)$ are shown in the left panels of Fig. 14. The profile likelihood scan is presented in the center panels. In the right panels, results are shown in the $\sigma_{t\bar{t}Z}$ versus $\sigma_{t\bar{t}W}$ plane. The 68% and 95% contours are obtained by sampling randomly from the fitted covariance matrix and extracting the contours which enclose 68.27% and 95.45% of the samples. We remove any assumptions about the energy scale of the NP made in Ref. [53] and report the ratio c_i/Λ^2. In cases where $\sigma_{SM+NP}(c_i)$ has the same minimum for all three processes, the profile likelihood is symmetric around this point, and we present results for $|c_i - c_{i,\text{min}}|$ to make this symmetry explicit.

Table 8: Observed best fit values for selected Wilson coefficients determined from this $t\bar{t}W$ and $t\bar{t}Z$ measurement, along with corresponding 68% and 95% CL intervals. In some cases the profile likelihood shows another local minimum that cannot be excluded; the number reported here is the global minimum.

Wilson coefficient	Best fit [TeV^{-2}]	68% CL [TeV^{-2}]	95% CL [TeV^{-2}]
$\bar{c}_{uW}/\Lambda^2$	1.7	$[-2.4, -0.5]$ and $[0.4, 2.4]$	$[-2.9, 2.9]$
$\lvert\bar{c}_H/\Lambda^2 - 16.8\,\text{TeV}^{-2}\rvert$	15.6	$[0, 23.0]$	$[0, 28.5]$
$\lvert\widetilde{c}_{3G}/\Lambda^2\rvert$	0.5	$[0, 0.7]$	$[0, 0.9]$
$\bar{c}_{3G}/\Lambda^2$	-0.4	$[-0.6, 0.1]$ and $[0.4, 0.7]$	$[-0.7, 1.0]$
$\bar{c}_{uG}/\Lambda^2$	0.2	$[0, 0.3]$	$[-1.0, -0.9]$ and $[-0.3, 0.4]$
$\lvert\bar{c}_{uB}/\Lambda^2\rvert$	1.6	$[0, 2.2]$	$[0, 2.7]$
$\bar{c}_{Hu}/\Lambda^2$	-9.3	$[-10.3, -8.0]$ and $[0, 2.1]$	$[-11.1, -6.5]$ and $[-1.6, 3.0]$
$\bar{c}_{2G}/\Lambda^2$	0.4	$[-0.9, -0.3]$ and $[-0.1, 0.6]$	$[-1.1, 0.8]$

9 Summary

A measurement of top quark pair production in association with a W or a Z boson using proton-proton collisions at 13 TeV is presented. The analysis is performed in the same-sign dilepton final state for $t\bar{t}W$, and the three- and four-lepton final states for $t\bar{t}Z$, and these three final states are used to extract the cross sections for $t\bar{t}W$ and $t\bar{t}Z$ production. For both processes the observed signal significance exceeds 5 standard deviations. The measured signal strength parameters are $1.23^{+0.19}_{-0.18}\,(\text{stat})^{+0.20}_{-0.18}\,(\text{syst})^{+0.13}_{-0.12}\,(\text{theo})$ and $1.17\ ^{+0.11}_{-0.10}\,(\text{stat})^{+0.14}_{-0.12}\,(\text{syst})^{+0.11}_{-0.12}\,(\text{theo})$ for $t\bar{t}W$ and $t\bar{t}Z$, respectively. The measured cross sections are $\sigma(t\bar{t}W) = 0.77^{+0.12}_{-0.11}\,(\text{stat})^{+0.13}_{-0.12}\,(\text{syst})\,\text{pb}$

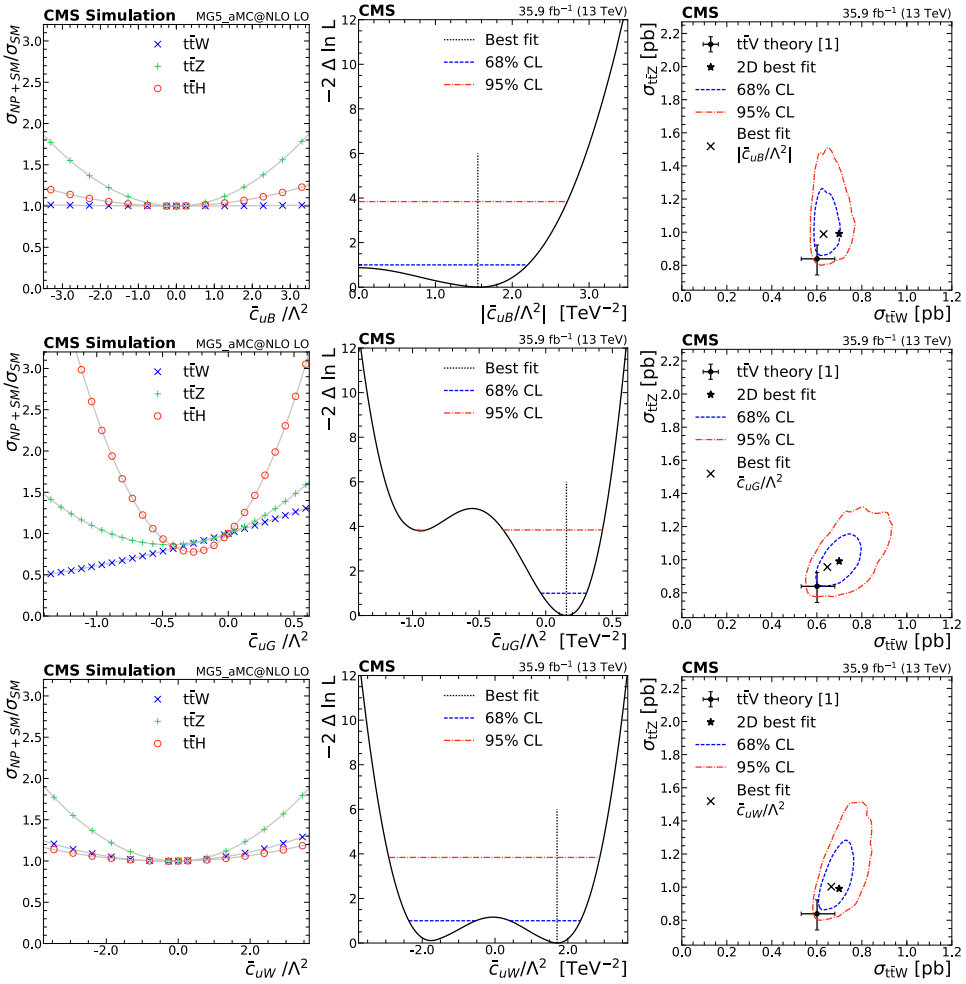


Figure 14: Left: signal strength as a function of selected Wilson coefficients for tt̄W (crosses), tt̄Z (pluses), and tt̄H (circles). Center: the 1D test statistic $q(c_i)$ scan as a function of c_i, profiling all other nuisance parameters. The global best fit value is indicated by a dotted line. Dashed and dash-dotted lines indicate 68% and 95% CL intervals, respectively. Right: The tt̄Z and tt̄W cross section corresponding to the global best fit c_i value is shown as a cross, along with the corresponding 68% (dashed) and 95% (dash-dotted) contours. The two-dimensional best fit to the tt̄W and tt̄Z cross sections is given by the star. The theory predictions [1] for tt̄W and tt̄Z are shown as a dot with bars representing their respective uncertainties.

and $\sigma(\mathrm{t\bar{t}Z}) = 0.99^{+0.09}_{-0.08}\,(\mathrm{stat})^{+0.12}_{-0.10}\,(\mathrm{syst})\,\mathrm{pb}$, in agreement with the standard model predictions. These results have been used to set constraints on the Wilson coefficients of dimension-six operators. Eight operators have been identified which are of particular interest because they change the expected cross sections of $\mathrm{t\bar{t}Z}$, $\mathrm{t\bar{t}W}$, or $\mathrm{t\bar{t}H}$ without significantly impacting expected background yields. Both $\mathrm{t\bar{t}Z}$ and $\mathrm{t\bar{t}H}$ are affected by $\mathcal{O}_{3G}$, $\mathcal{O}_{\widetilde{3G}}$, $\mathcal{O}_{2G}$, and $\mathcal{O}_{uB}$. Only $\mathrm{t\bar{t}Z}$ is affected by $\mathcal{O}_{Hu}$, while $\mathcal{O}_{H}$ affects only $\mathrm{t\bar{t}H}$. All three processes $\mathrm{t\bar{t}Z}$, $\mathrm{t\bar{t}W}$, and $\mathrm{t\bar{t}H}$ are affected by $\mathcal{O}_{uG}$ and $\mathcal{O}_{uW}$. In cases where new physics beyond the standard model modifies the expected $\mathrm{t\bar{t}Z}$ cross section, the sensitivity is mainly determined by $\mathrm{t\bar{t}Z}$ and the fit is able to match the observed excess in data. No operators were identified which provide an independent handle on $\mathrm{t\bar{t}W}$. The constraints presented, obtained by considering one operator at a time, are a useful first step toward more global approaches.

Acknowledgments

We congratulate our colleagues in the CERN accelerator departments for the excellent performance of the LHC and thank the technical and administrative staffs at CERN and at other CMS institutes for their contributions to the success of the CMS effort. In addition, we gratefully acknowledge the computing centers and personnel of the Worldwide LHC Computing Grid for delivering so effectively the computing infrastructure essential to our analyses. Finally, we acknowledge the enduring support for the construction and operation of the LHC and the CMS detector provided by the following funding agencies: BMWFW and FWF (Austria); FNRS and FWO (Belgium); CNPq, CAPES, FAPERJ, and FAPESP (Brazil); MES (Bulgaria); CERN; CAS, MoST, and NSFC (China); COLCIENCIAS (Colombia); MSES and CSF (Croatia); RPF (Cyprus); SENESCYT (Ecuador); MoER, ERC IUT, and ERDF (Estonia); Academy of Finland, MEC, and HIP (Finland); CEA and CNRS/IN2P3 (France); BMBF, DFG, and HGF (Germany); GSRT (Greece); OTKA and NIH (Hungary); DAE and DST (India); IPM (Iran); SFI (Ireland); INFN (Italy); MSIP and NRF (Republic of Korea); LAS (Lithuania); MOE and UM (Malaysia); BUAP, CINVESTAV, CONACYT, LNS, SEP, and UASLP-FAI (Mexico); MBIE (New Zealand); PAEC (Pakistan); MSHE and NSC (Poland); FCT (Portugal); JINR (Dubna); MON, RosAtom, RAS, RFBR and RAEP (Russia); MESTD (Serbia); SEIDI, CPAN, PCTI and FEDER (Spain); Swiss Funding Agencies (Switzerland); MST (Taipei); ThEPCenter, IPST, STAR, and NSTDA (Thailand); TUBITAK and TAEK (Turkey); NASU and SFFR (Ukraine); STFC (United Kingdom); DOE and NSF (USA).

Individuals have received support from the Marie-Curie program and the European Research Council and Horizon 2020 Grant, contract No. 675440 (European Union); the Leventis Foundation; the A. P. Sloan Foundation; the Alexander von Humboldt Foundation; the Belgian Federal Science Policy Office; the Fonds pour la Formation à la Recherche dans l'Industrie et dans l'Agriculture (FRIA-Belgium); the Agentschap voor Innovatie door Wetenschap en Technologie (IWT-Belgium); the Ministry of Education, Youth and Sports (MEYS) of the Czech Republic; the Council of Science and Industrial Research, India; the HOMING PLUS program of the Foundation for Polish Science, cofinanced from European Union, Regional Development Fund, the Mobility Plus program of the Ministry of Science and Higher Education, the National Science Center (Poland), contracts Harmonia 2014/14/M/ST2/00428, Opus 2014/13/B/ST2/02543, 2014/15/B/ST2/03998, and 2015/19/B/ST2/02861, Sonata-bis 2012/07/E/ST2/01406; the National Priorities Research Program by Qatar National Research Fund; the Programa Severo Ochoa del Principado de Asturias; the Thalis and Aristeia programs cofinanced by EU-ESF and the Greek NSRF; the Rachadapisek Sompot Fund for Postdoctoral Fellowship, Chulalongkorn University and the Chulalongkorn Academic into Its 2nd Century Project Advancement Project (Thailand); the Welch Foundation, contract C-1845; and the Weston Havens Foundation (USA).

References

[1] D. de Florian et al., "Handbook of LHC Higgs cross sections: 4. deciphering the nature of the Higgs sector", CERN Report CERN-2017-002-M, 2016. doi:10.23731/CYRM-2017-002, arXiv:1610.07922.

[2] O. Bessidskaia Bylund et al., "Probing top quark neutral couplings in the Standard Model Effective Field Theory at NLO in QCD", *JHEP* **05** (2016) 052, doi:10.1007/JHEP05(2016)052, arXiv:1601.08193.

[3] C. Englert, R. Kogler, H. Schulz, and M. Spannowsky, "Higgs coupling measurements at the LHC", *Eur. Phys. J. C* **76** (2016) 393, doi:10.1140/epjc/s10052-016-4227-1, arXiv:1511.05170.

[4] CMS Collaboration, "Measurement of associated production of vector bosons and $t\bar{t}$ in pp collisions at $\sqrt{s} = 7$ TeV", *Phys. Rev. Lett.* **110** (2013) 172002, doi:10.1103/PhysRevLett.110.172002, arXiv:1303.3239.

[5] CMS Collaboration, "Measurement of top quark-antiquark pair production in association with a W or Z boson in pp collisions at $\sqrt{s} = 8$ TeV", *Eur. Phys. J. C* **74** (2014) 3060, doi:10.1140/epjc/s10052-014-3060-7, arXiv:1406.7830.

[6] CMS Collaboration, "Observation of top quark pairs produced in association with a vector boson in pp collisions at $\sqrt{s} = 8$ TeV", *JHEP* **01** (2016) 096, doi:10.1007/JHEP01(2016)096, arXiv:1510.01131.

[7] ATLAS Collaboration, "Measurement of the $t\bar{t}W$ and $t\bar{t}Z$ production cross sections in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector", *JHEP* **11** (2015) 172, doi:10.1007/JHEP11(2015)172, arXiv:1509.05276.

[8] ATLAS Collaboration, "Measurement of the $t\bar{t}Z$ and $t\bar{t}W$ production cross sections in multilepton final states using 3.2 fb^{-1} of pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector", *Eur. Phys. J. C* **77** (2017) 40, doi:10.1140/epjc/s10052-016-4574-y, arXiv:1609.01599.

[9] K. G. Wilson, "Non-lagrangian models of current algebra", *Phys. Rev.* **179** (1969) 1499, doi:10.1103/PhysRev.179.1499.

[10] CMS Collaboration, "The CMS experiment at the CERN LHC", *JINST* **3** (2008) S08004, doi:10.1088/1748-0221/3/08/S08004.

[11] CMS Collaboration, "The CMS trigger system", *JINST* **12** (2017) P01020, doi:10.1088/1748-0221/12/01/P01020, arXiv:1609.02366.

[12] J. Alwall et al., "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations", *JHEP* **07** (2014) 079, doi:10.1007/JHEP07(2014)079, arXiv:1405.0301.

[13] S. Alioli, P. Nason, C. Oleari, and E. Re, "A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX", *JHEP* **06** (2010) 043, doi:10.1007/JHEP06(2010)043, arXiv:1002.2581.

[14] H. B. Hartanto, B. Jager, L. Reina, and D. Wackeroth, "Higgs boson production in association with top quarks in the POWHEG BOX", *Phys. Rev. D* **91** (2015) 094003, doi:10.1103/PhysRevD.91.094003, arXiv:1501.04498.

[15] T. Melia, P. Nason, R. Rontsch, and G. Zanderighi, "W^+W^-, WZ and ZZ production in the POWHEG BOX", *JHEP* **11** (2011) 078, doi:10.1007/JHEP11(2011)078, arXiv:1107.5051.

[16] P. Nason and G. Zanderighi, "W^+W^- , WZ and ZZ production in the POWHEG-BOX-V2", *Eur. Phys. J. C* **74** (2014) 2702, doi:10.1140/epjc/s10052-013-2702-5, arXiv:1311.1365.

[17] J. M. Campbell and R. K. Ellis, "MCFM for the Tevatron and the LHC", *Nucl. Phys. Proc. Suppl.* **205-206** (2010) 10, doi:10.1016/j.nuclphysbps.2010.08.011, arXiv:1007.3492.

[18] F. Cascioli et al., "ZZ production at hadron colliders in NNLO QCD", *Phys. Lett. B* **735** (2014) 311, doi:10.1016/j.physletb.2014.06.056, arXiv:1405.2219.

[19] F. Caola, K. Melnikov, R. Röntsch, and L. Tancredi, "QCD corrections to ZZ production in gluon fusion at the LHC", *Phys. Rev. D* **92** (2015) 094028, doi:10.1103/PhysRevD.92.094028, arXiv:1509.06734.

[20] NNPDF Collaboration, "Parton distributions for the LHC Run II", *JHEP* **04** (2015) 040, doi:10.1007/JHEP04(2015)040, arXiv:1410.8849.

[21] T. Sjöstrand, S. Mrenna, and P. Z. Skands, "A brief introduction to PYTHIA 8.1", *Comput. Phys. Commun.* **178** (2008) 852, doi:10.1016/j.cpc.2008.01.036, arXiv:0710.3820.

[22] T. Sjöstrand et al., "An introduction to PYTHIA 8.2", *Comput. Phys. Commun.* **191** (2015) 159, doi:10.1016/j.cpc.2015.01.024, arXiv:1410.3012.

[23] P. Skands, S. Carrazza, and J. Rojo, "Tuning PYTHIA 8.1: the Monash 2013 tune", *Eur. Phys. J. C* **74** (2014) 3024, doi:10.1140/epjc/s10052-014-3024-y, arXiv:1404.5630.

[24] CMS Collaboration, "Event generator tunes obtained from underlying event and multiparton scattering measurements", *Eur. Phys. J. C* **76** (2016) 155, doi:10.1140/epjc/s10052-016-3988-x, arXiv:1512.00815.

[25] J. Alwall et al., "Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions", *Eur. Phys. J. C* **53** (2008) 473, doi:10.1140/epjc/s10052-007-0490-5, arXiv:0706.2569.

[26] R. Frederix and S. Frixione, "Merging meets matching in MC@NLO", *JHEP* **12** (2012) 061, doi:10.1007/JHEP12(2012)061, arXiv:1209.6215.

[27] GEANT4 Collaboration, "GEANT4 — a simulation toolkit", *Nucl. Instrum. Meth. A* **506** (2003) 250, doi:10.1016/S0168-9002(03)01368-8.

[28] CMS Collaboration, "Particle-flow reconstruction and global event description with the CMS detector", *JINST* **12** (2017) P10003, doi:10.1088/1748-0221/12/10/P10003, arXiv:1706.04965.

[29] M. Cacciari, G. P. Salam, and G. Soyez, "The anti-k_t jet clustering algorithm", *JHEP* **04** (2008) 063, doi:10.1088/1126-6708/2008/04/063, arXiv:0802.1189.

[30] M. Cacciari, G. P. Salam, and G. Soyez, "FastJet user manual", *Eur. Phys. J. C* **72** (2012) 1896, doi:10.1140/epjc/s10052-012-1896-2, arXiv:1111.6097.

[31] M. Cacciari and G. P. Salam, "Dispelling the N^3 myth for the k_t jet-finder", *Phys. Lett. B* **641** (2006) 57, doi:10.1016/j.physletb.2006.08.037, arXiv:hep-ph/0512210.

[32] CMS Collaboration, "Pileup removal algorithms", CMS Physics Analysis Summary CMS-PAS-JME-14-001, 2014.

[33] CMS Collaboration, "Identification of b-quark jets with the CMS experiment", *JINST* **8** (2013) P04013, doi:10.1088/1748-0221/8/04/P04013, arXiv:1211.4462.

[34] CMS Collaboration, "Identification of b quark jets at the CMS experiment in the LHC Run 2", CMS Physics Analysis Summary CMS-PAS-BTV-15-001, 2016.

[35] Particle Data Group, "Review of particle physics", *Chin. Phys. C* **40** (2016) 100001, doi:10.1088/1674-1137/40/10/100001.

[36] H. Voss, A. Höcker, J. Stelzer, and F. Tegenfeldt, "TMVA, the toolkit for multivariate data analysis with ROOT", in *XIth International Workshop on Advanced Computing and Analysis Techniques in Physics Research (ACAT)*, p. 40. 2007. arXiv:physics/0703039.

[37] CMS Collaboration, "Search for new physics in same-sign dilepton events in proton-proton collisions at $\sqrt{s} = 13$ TeV", *Eur. Phys. J. C* **76** (2016) 439, doi:10.1140/epjc/s10052-016-4261-z, arXiv:1605.03171.

[38] CMS Collaboration, "Search for physics beyond the standard model in events with two leptons of same sign, missing transverse momentum, and jets in proton-proton collisions at $\sqrt{s} = 13$ TeV", *Eur. Phys. J. C* **77** (2017) 578, doi:10.1140/epjc/s10052-017-5079-z, arXiv:1704.07323.

[39] J. Campbell, R. K. Ellis, and R. Röntsch, "Single top production in association with a Z boson at the LHC", *Phys. Rev. D* **87** (2013) 114006, doi:10.1103/PhysRevD.87.114006, arXiv:1302.3856.

[40] S. Frixione et al., "Electroweak and QCD corrections to top-pair hadroproduction in association with heavy bosons", *JHEP* **06** (2015) 184, doi:10.1007/JHEP06(2015)184, arXiv:1504.03446.

[41] CMS Collaboration, "CMS luminosity measurement for the 2016 data taking period", CMS Physics Analysis Summary CMS-PAS-LUM-17-001, 2017.

[42] ATLAS Collaboration, "Measurement of the inelastic proton-proton cross section at $\sqrt{s} = 13$ TeV with the ATLAS detector at the LHC", *Phys. Rev. Lett.* **117** (2016) 182002, doi:10.1103/PhysRevLett.117.182002, arXiv:1606.02625.

[43] CMS Collaboration, "Performance of CMS muon reconstruction in pp collision events at $\sqrt{s} = 7$ TeV", *JINST* **7** (2012) P10002, doi:10.1088/1748-0221/7/10/P10002, arXiv:1206.4071.

[44] CMS Collaboration, "Performance of electron reconstruction and selection with the CMS detector in proton-proton collisions at $\sqrt{s} = 8$ TeV", *JINST* **10** (2015) P06005, doi:10.1088/1748-0221/10/06/P06005, arXiv:1502.02701.

[45] CMS Collaboration, "Performance of the CMS missing transverse momentum reconstruction in pp data at $\sqrt{s} = 8$ TeV", *JINST* **10** (2015) P02006, doi:10.1088/1748-0221/10/02/P02006, arXiv:1411.0511.

[46] CMS Collaboration, "Performance of missing energy reconstruction in 13 TeV pp collision data using the CMS detector", CMS Physics Analysis Summary CMS-PAS-JME-16-004, 2016.

[47] T. Junk, "Confidence level computation for combining searches with small statistics", *Nucl. Instrum. Meth. A* **434** (1999) 435, doi:10.1016/S0168-9002(99)00498-2, arXiv:hep-ex/9902006.

[48] A. L. Read, "Presentation of search results: the CL_s technique", in *Durham IPPP Workshop: Advanced Statistical Techniques in Particle Physics*, p. 2693. Durham, UK, March, 2002. [J. Phys. G 28 (2002) 2693]. doi:10.1088/0954-3899/28/10/313.

[49] ATLAS and CMS Collaborations, "Procedure for the LHC Higgs boson search combination in summer 2011", ATL-PHYS-PUB-2011-011, CMS NOTE-2011/005, 2011.

[50] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, "Asymptotic formulae for likelihood-based tests of new physics", *Eur. Phys. J. C* **71** (2011) 1554, doi:10.1140/epjc/s10052-011-1554-0, arXiv:1007.1727. [Erratum: doi:10.1140/epjc/s10052-013-2501-z].

[51] W. Buchmüller and D. Wyler, "Effective lagrangian analysis of new interactions and flavour conservation", *Nucl. Phys. B* **268** (1986) 621, doi:10.1016/0550-3213(86)90262-2.

[52] B. Grzadkowski, M. Iskrzynski, M. Misiak, and J. Rosiek, "Dimension-six terms in the Standard Model Lagrangian", *JHEP* **10** (2010) 085, doi:10.1007/JHEP10(2010)085, arXiv:1008.4884.

[53] A. Alloul, B. Fuks, and V. Sanz, "Phenomenology of the Higgs effective Lagrangian via FeynRules", *JHEP* **04** (2014) 110, doi:10.1007/JHEP04(2014)110, arXiv:1310.5150.

[54] J. Ellis, V. Sanz, and T. You, "Complete Higgs sector constraints on dimension-6 operators", *JHEP* **07** (2014) 036, doi:10.1007/JHEP07(2014)036, arXiv:1404.3667.

[55] K. Whisnant, J.-M. Yang, B.-L. Young, and X. Zhang, "Dimension-six CP-conserving operators of the third-family quarks and their effects on collider observables", *Phys. Rev. D* **56** (1997) 467, doi:10.1103/PhysRevD.56.467, arXiv:hep-ph/9702305.

[56] E. L. Berger, Q.-H. Cao, and I. Low, "Model independent constraints among the Wtb, Zb$\bar{\text{b}}$, and Zt$\bar{\text{t}}$ couplings", *Phys. Rev. D* **80** (2009) 074020, doi:10.1103/PhysRevD.80.074020, arXiv:0907.2191.

[57] R. Röntsch and M. Schulze, "Constraining couplings of top quarks to the Z boson in t$\bar{\text{t}}$+Z production at the LHC", *JHEP* **07** (2014) 091, doi:10.1007/JHEP09(2015)132, arXiv:1404.1005. [Erratum: doi:10.1007/JHEP07(2014)091].

[58] E. Malkawi and C. P. Yuan, "Global analysis of the top quark couplings to gauge bosons", *Phys. Rev. D* **50** (1994) 4462, doi:10.1103/PhysRevD.50.4462, arXiv:hep-ph/9405322.

28

[59] C. Zhang, N. Greiner, and S. Willenbrock, "Constraints on nonstandard top quark couplings", *Phys. Rev. D* **86** (2012) 014024, doi:10.1103/PhysRevD.86.014024, arXiv:1201.6670.

[60] A. Tonero and R. Rosenfeld, "Dipole-induced anomalous top quark couplings at the LHC", *Phys. Rev. D* **90** (2014) 017701, doi:10.1103/PhysRevD.90.017701, arXiv:1404.2581.

[61] C. Degrande et al., "Effective field theory: A modern approach to anomalous couplings", *Annals Phys.* **335** (2013) 21, doi:10.1016/j.aop.2013.04.016, arXiv:1205.4231.

[62] A. Alloul et al., "FEYNRULES 2.0 — A complete toolbox for tree-level phenomenology", *Comput. Phys. Commun.* **185** (2014) 2250, doi:10.1016/j.cpc.2014.04.012, arXiv:1310.1921.

[63] S. Frixione and B. R. Webber, "Matching NLO QCD computations and parton shower simulations", *JHEP* **06** (2002) 029, doi:10.1088/1126-6708/2002/06/029, arXiv:hep-ph/0204244.

[64] J. Ellis, "TikZ-Feynman: Feynman diagrams with TikZ", *Comput. Phys. Commun.* **210** (2017) 103, doi:10.1016/j.cpc.2016.08.019, arXiv:1601.05437.

7.1 TRIGGER AND OBJECT SELECTION

Similar to the analysis of the SUSY search in the multi-lepton channel, described in Chapter 6, an optimized lepton selection criterion is applied in this analysis to reduce the background from nonprompt leptons while keeping a high signal efficiency. Two sets of selection criteria are applied: tight lepton selection is applied to select signal-like leptons while loose selection criterion is applied to select a sample of events which is then used to estimate the nonprompt lepton background. A detailed description of the individual selection criteria for electrons and muons is shown in the Table 7.1.

Electrons		
Criterion	Loose	Tight
$p_{\mathrm{T}} > 10$ GeV	✓	✓
$\lvert\eta\rvert < 2.5$	✓	✓
$\lvert d_{xy}\rvert < 0.05$ cm	✓	✓
$\lvert d_{z}\rvert < 0.1$ cm	✓	✓
$\mathrm{SIP}_{3D} < 4$	✓	✓
conversion rejection	✓	✓
no missing pixel hits	✓	✓
charge consistency	✓	✓
electron MVA	95% eff.	90% eff.
I_{rel} EA corr.	$< 1.$	< 0.1

Muons		
Criterion	Loose	Tight
$p_{\mathrm{T}} > 10$ GeV	✓	✓
$\lvert\eta\rvert < 2.4$	✓	✓
$\lvert d_{xy}\rvert < 0.05$ cm	✓	✓
$\lvert d_{z}\rvert < 0.1$ cm	✓	✓
$\mathrm{SIP}_{3D} < 4$	✓	✓
Medium ID	✓	✓
I_{rel} $\Delta\beta$ corr.	< 0.6	< 0.15

Table 7.1: List of all quality criteria for electrons and muons for passing the loose and the tight selection in the $\mathrm{t\bar{t}W}$ analysis.

Jets that pass the loose identification criterion, defined in Subsection 5.2.4, are used in this analysis. Among these jets, the ones with $p_{\mathrm{T}} > 30$ GeV and within $|\eta| < 2.4$ are selected. The CSVv2 b tagging algorithm is used to identify b jets in the analysis. Similar to the SUSY analysis jets lying within $\Delta R = 0.4$ of lepton's vicinity are rejected.

Events are selected by online triggers that require the presence of at least one electron or muon. The unprescaled single lepton triggers with lowest p_{T} threshold that cover full η ranges for electrons and muons are HLT_Iso(Tk)Mu24 and HLT_Ele27_WPTight_Gsf. From the names of the triggers, the p_{T} thresholds applied on the leptons, $p_{\mathrm{T}} > 24$ for muons and $p_{\mathrm{T}} > 27$ for electrons, can be deduced. These triggers are chosen to give flexibility for the second lepton to be free of any biased selection at the trigger level.

In order to take into account all possible differences in the trigger efficiency between data and simulation, the efficiency of the trigger selection is studied in each dilepton channel. Firstly, single lepton trigger efficiencies are measured in simulated W+jets events with exactly one lepton that passes the identification and isolation requirement. These efficiencies are presented in Fig. 7.2 as a function of the lepton p_{T} and $|\eta|$ for muons and electrons. The same figures show the measured efficiencies in the unbiased data sample, i.e. the data set selected with triggers that require large p_{T}^{miss}.

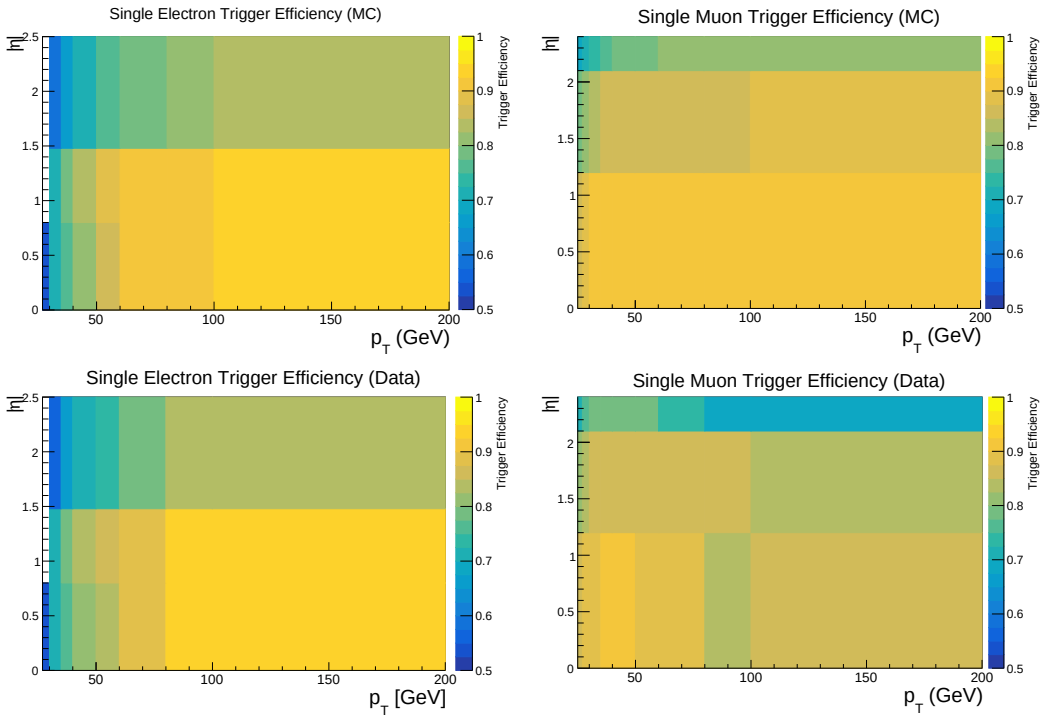


Figure 7.2: Measured efficiencies for single-electron(left) and single-muon (right) triggers as a function of lepton p_{T} and η measured in MC (top) and data (bottom).

The efficiency of dilepton event selection with the single lepton triggers is measured according to the following formula:

$$\text{Total Eff} = 1 - [(1 - \varepsilon_1(p_{\mathrm{T}}, \eta)) \times (1 - \varepsilon_2(p_{\mathrm{T}}, \eta))],$$

where $\varepsilon_i(p_{\mathrm{T}}, \eta)$ is taken from the efficiency maps in Fig. 7.2 either from the data or from the simulation. The procedure applied here assumes that the trigger efficiencies for each lepton in the event are factorisable. To validate this assumption we compare these efficiencies to the single lepton trigger efficiency measured in dilepton events in $\mathrm{t\bar{t}W}$ MC sample and dilepton events in the data set selected with p_{T}^{miss} triggers. These efficiencies are shown in Fig. 7.3. The level of agreement between red and blue points shows the goodness of the assumption on the factorization. In general, good agreement was found between all three efficiencies. Corrections are needed to apply in the high p_{T} region in case of dielectron events and in all p_{T} spectra of dimuon events. Overall the efficiency of selecting dilepton events with single lepton triggers is greater than 92%.

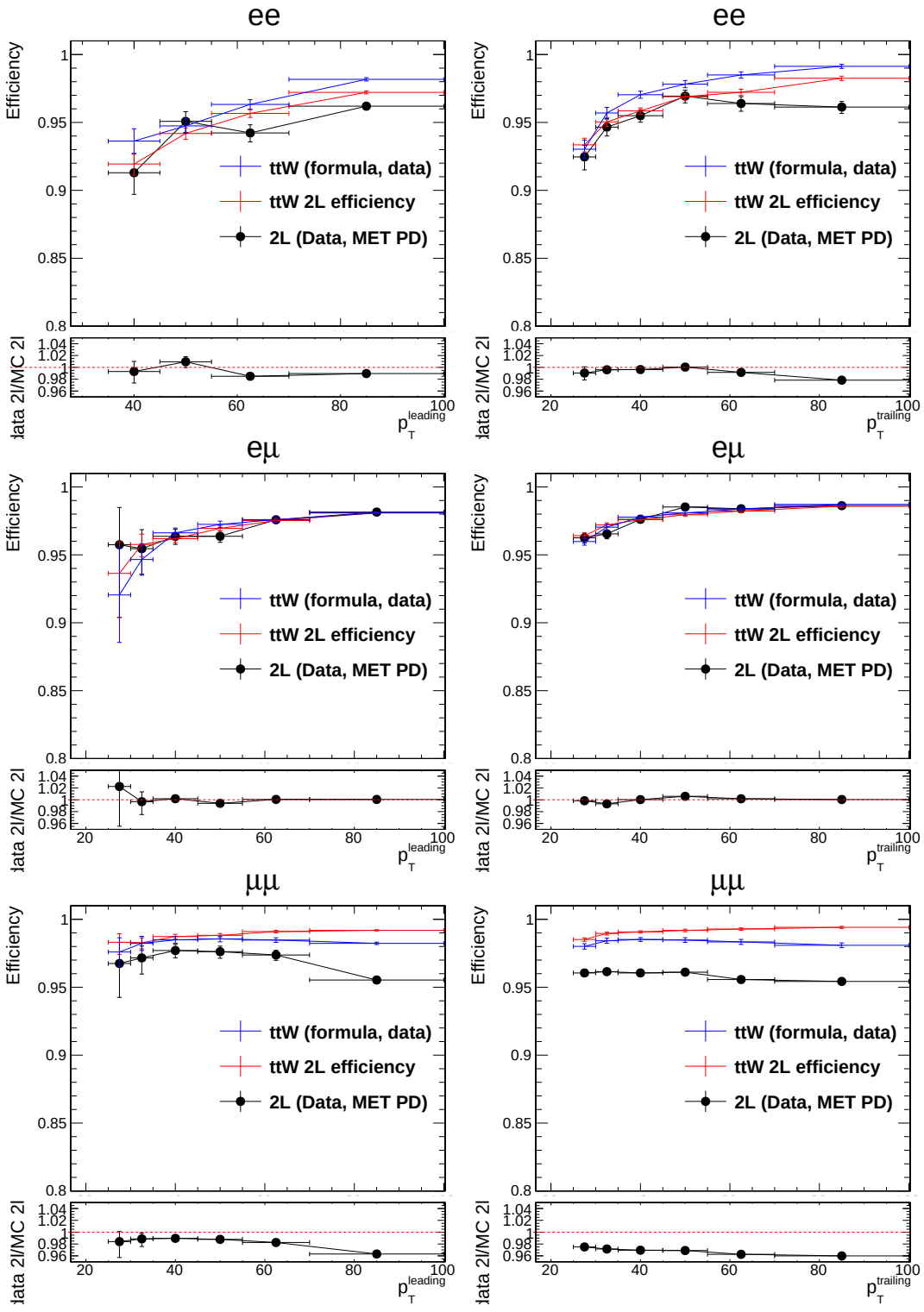


Figure 7.3: Trigger efficiencies for ee (top), $e\mu$ (middle) and $\mu\mu$ (bottom)-channels as a function of the leading (left) and trailing (right) lepton p_{T} calculated directly in dilepton events in $\mathrm{t\bar{t}W}$ MC and events selected with p_{T}^{miss} triggers. The single lepton trigger efficiencies for dilepton event selection efficiency estimation are measured in the data set selected with triggers that require large p_{T}^{miss}.

7.2 STRATEGY

We select events that contain exactly two same-charge leptons ($\mu\mu$, μe, ee) where the leading and trailing lepton transverse momenta are required to be above 25 GeV (greater than 27 GeV in case of electrons); to avoid inefficiencies due to the trigger selection in the dielectron channel, the leading electron is required to have $p_{\mathrm{T}} > 40$ GeV. Events containing additional leptons passing loose lepton identification requirement are vetoed. The $\mathrm{t\bar{t}W}$ process in the same-charge dilepton final state gives exactly the same number of light and heavy flavor jets as the $\mathrm{t\bar{t}}$ process in the semi-leptonic final state:

$$pp \to t\bar{t}W \to (t \to b\ell^{\pm}\nu)(\bar{t} \to \bar{b}jj)(W \to \ell^{\pm}\nu),$$

$$pp \to t\bar{t} \to (t \to b\ell^{\pm}\nu)(\bar{t} \to \bar{b}jj).$$

An additional lepton in the latter process can arise from a semileptonic decay of a b quark for instance. In general, we expect that subsequent reconstructed jets will contain the lepton. Discarding jets that are close to the leptons, the distributions of jets and b-tagged jets multiplicity will be shifted to lower values with respect to the same distribution in the $\mathrm{t\bar{t}W}$ process. Therefore, to suppress the nonprompt lepton background and to increase signal-over-background ratio events with at least 2 jets and 1 b-tagged jet are considered in the analysis. This requirement reduces the contribution from $\mathrm{t\bar{t}}$ in the same-charge dilepton category by 50%, while for the $\mathrm{t\bar{t}W}$ process, only 30% of events are discarded.

Another process that has a high cross section and gives two leptons in the final state is the DY process. This process can pass the selection with a same-charge pair through the mismeasurement of the electron charge in the CMS detector. The charge misidentification rate is quite significant in the regions, where the material budget of the detector is high [95]. It should also be mentioned that the charge misidentification rate for muons is usually 100 times lower than for electrons, therefore muons give a marginal contribution in this type of background. To suppress the contribution from a dielectron pair with the misidentified charge of one of the electrons, the invariant mass of two electrons is required to lie outside a 15 GeV window around the Z pole mass. Still misidentification of electron charge can play a significant role in a case when Z/γ^* boson decays to a pair of τ leptons with subsequent decay either to 2 electrons or an electron-muon pair. This remaining contribution to the background from the DY process is suppressed by requiring p_{T}^{miss} to be greater than 30 GeV. In addition to all these requirements, the invariant mass of the lepton pair is required to be greater than 12 GeV in order to suppress the contribution from quarkonium processes.

7.2.1 *Multivariate analysis*

After applying these requirements, the remaining background is still quite high with respect to the signal yields. Therefore, a multivariate analysis (MVA) has been developed. Among the kinematic variables that are examined as inputs to the multivariate analysis training, the following ones are found to provide the best discrimination between signal and background: N_{j}, N_{b}, H_{T}, the highest-p_{T} and the second highest-p_{T} jets, the leading and the trailing lepton p_{T}, p_{T}^{miss}, M_{T} for each of the leptons and ΔR between the trailing lepton and the closest selected jet. The discrimination power of these variables is demonstrated in Fig. 7.4 and 7.5.

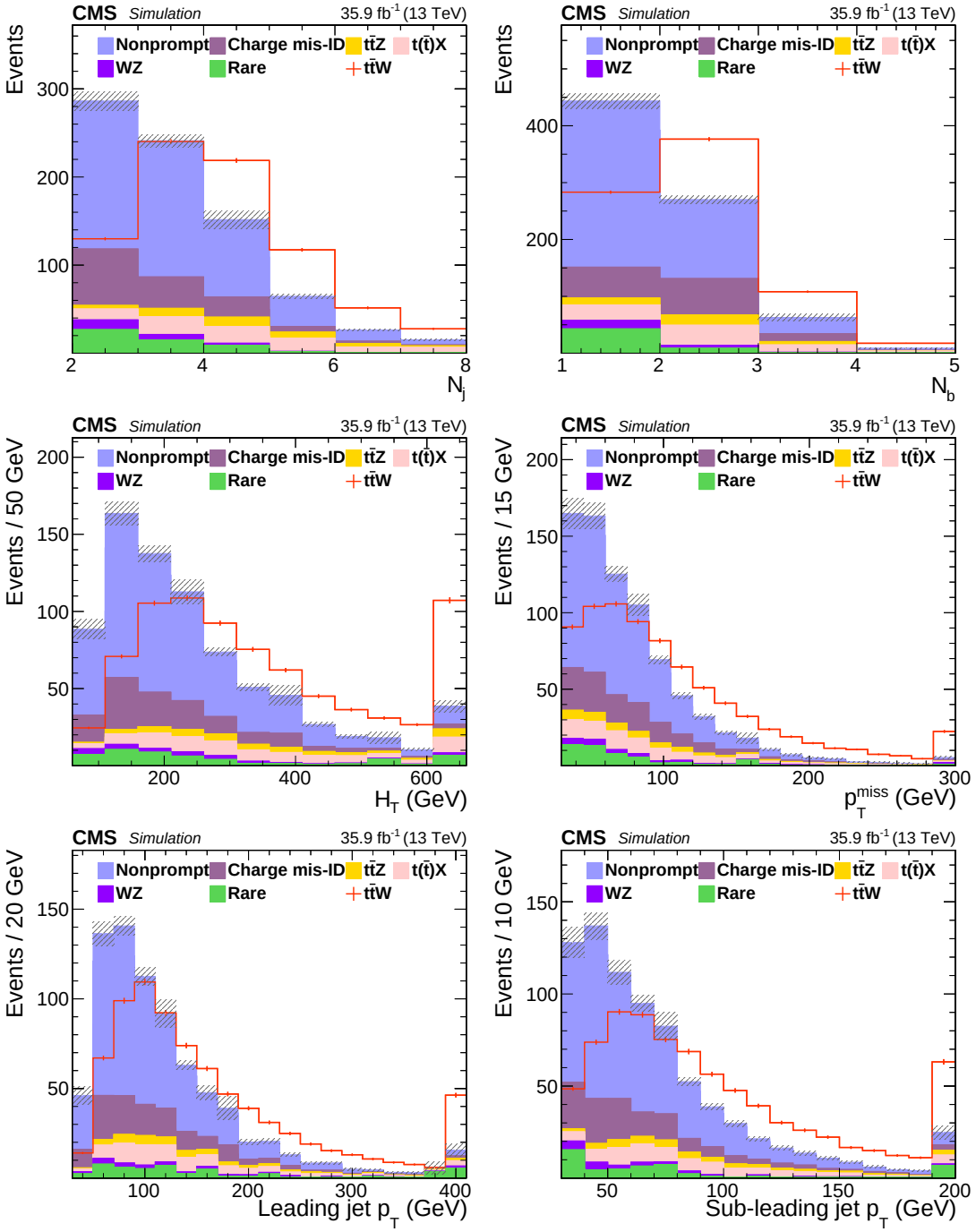


Figure 7.4: Distributions of kinematical variables used for the training in MVA. Number of events for $t\bar{t}W$ is scaled to integral number of events in background. Total background is scaled to the integrated luminosity of 35.9 fb^{-1} collected in 2016.

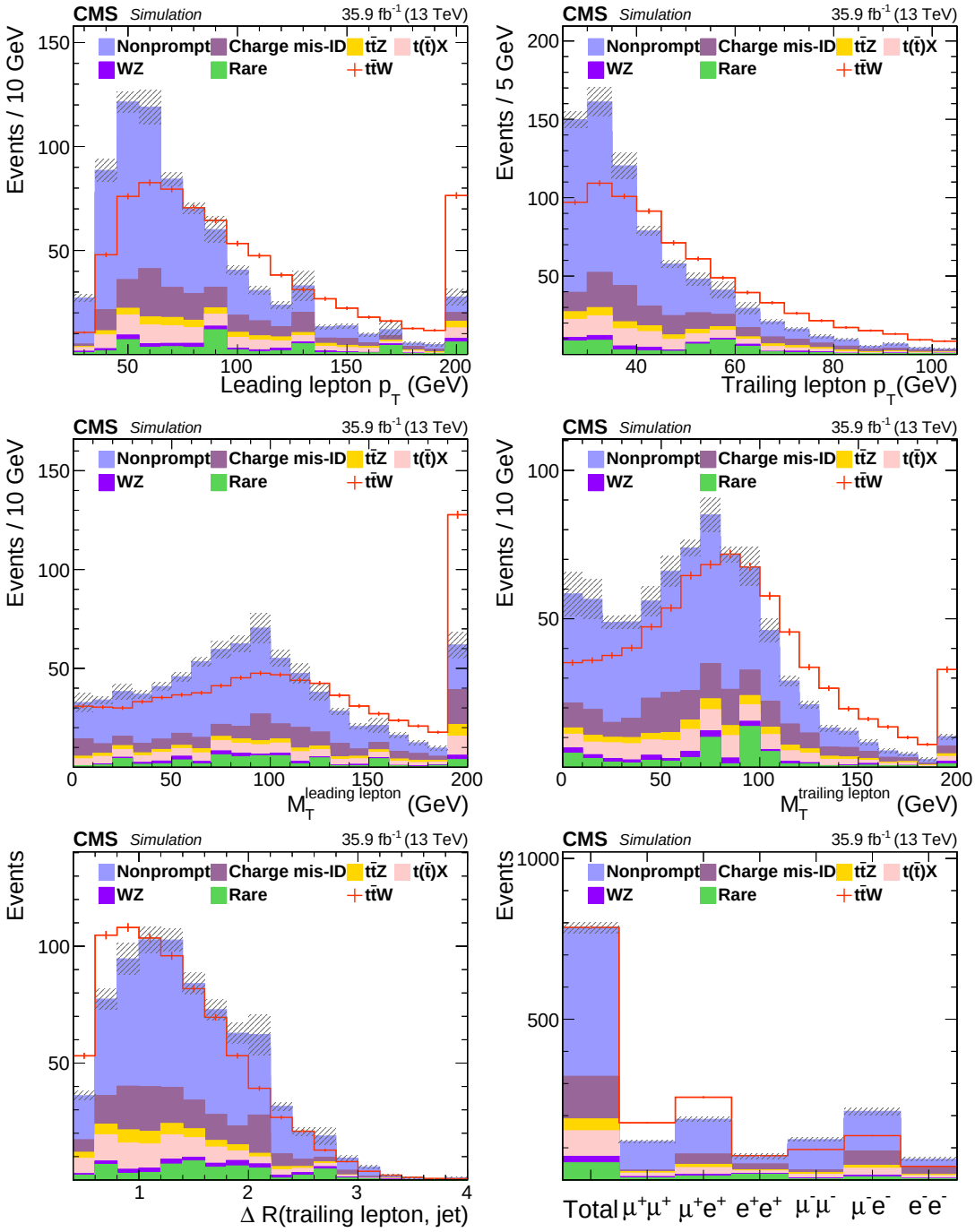


Figure 7.5: Distributions of kinematical variables used for the training in MVA. Additionally, flavor composition is added. Number of events for $t\bar{t}W$ is scaled to integral number of events in background. Total background is scaled to the integrated luminosity of 35.9 fb^{-1} collected in 2016.

A decision tree classifier [104] with a gradient boost is used. The training is performed by using the $t\bar{t}W$ simulation as signal and the $t\bar{t}$ simulation as background. Events are equally split for training and testing. The settings of the BDT are chosen to give the optimal results: the BDT uses a forest of 1000 trees, with a maximum depth of 2 nodes. Each node needs to contain at least 2.5% of all events. The learning rate is taken as 0.1 and on top of it bagged boosting with a random sample splitting of 50% is used. The BDT gives a score that ranges from -1 for background-like events to 1 for signal-like events.

In Fig. 7.6 the discriminating variable delivered by the MVA is shown. The MVA helps to increase signal-over-background ratio from 0.17, obtained with the pre-selected selection, to 0.7 in the last bin of the MVA distribution. Nevertheless, all events obtained with the pre-selection criterion are used in the analysis. The BDT output is split into three categories: BDT value < 0, $0 <$ BDT < 0.6 and BDT > 0.6. The first region is completely dominated in the nonprompt lepton background and is used in the analysis as a control region in data. In the latter two categories, N_j and N_b are used to form five exclusive categories: for events with N_j equal to 3 and > 3 we additionally split in N_b equal to 1 and > 1, while for events with $N_j = 2$ no further split is needed. Furthermore, we would like to benefit from asymmetric production in the $t\bar{t}W$ process, therefore, we split events into 2 positive and 2 negative charged leptons categories. In low BDT value categories we split events according to the number of jets, namely 2, 3 and > 3. In total, 23 signal regions are used to extract the cross section for the $t\bar{t}W$ process.

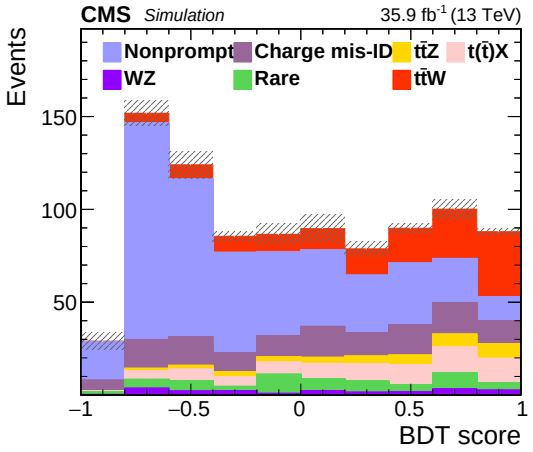


Figure 7.6: Distribution of the boosted decision tree classifier for background and signal processes after applying the pre-selected selection criteria. The expected contribution from the different background processes estimated from simulation and scaled to the integrated luminosity of 35.9 fb^{-1} is shown. The shaded band represents the statistical uncertainty derived from MC samples.

7.3 BACKGROUND ESTIMATION

Similar to the SUSY analysis, the background processes can be split into a few categories. Nonprompt lepton background and background induced by the mismeasurement of the lepton charge are the most significant, so dedicated data-driven techniques are used for their estimation. The remaining categories consist of either the backgrounds, that produces more than 2 leptons in the final state, but one of the lepton falls out of detector's acceptance or does not pass the selection criteria, or the processes that

produce exactly the same signature as the $t\bar{t}W$ process in same-charge dilepton final state. The largest contribution in this category comes from the WZ, $t\bar{t}Z$, $t\bar{t}H$ and single top production with a BEH boson (tHq) processes. These processes along with the other remaining processes are grouped into the rare SM processes category, and are estimated using MC simulation.

7.3.1 *Nonprompt lepton background*

The method used for the estimation of the nonprompt lepton background does not differ from the one described in Section 6.4.1. Similarly, here we measure the misidentification rate as a function of the corrected lepton p_T and η, where the corrected p_T is defined according to Eq. 6.2 with a substitution of I_{rel} instead of I_{mini}. The values of I_{rel} and the electron MVA for loose identification criteria are adjusted to have a low misidentification rate. Thus, good statistical precision for the nonprompt lepton background estimation could be achieved.

The misidentification rates as a function of corrected lepton p_T and η are shown in Fig. 7.7. A similar event selection, triggers and method for the electroweak scale factor calculation are used as discussed in the Section 6.4.1. In general, good agreement between observed misidentification rates measured in data and simulation was found, apart from the high lepton p_T category in the data. This region is dominated by the prompt leptons from W+jets and DY processes, and, thus, it is not fully reliable for the nonprompt lepton background estimation. Therefore, for leptons with corrected lepton p_T greater than 65 GeV, the misidentification rate from the 45 GeV $< p_T <$ 65 GeV category is applied.

The tight-to-loose method for the $t\bar{t}W$ analysis is verified in the closure test performed with the MC $t\bar{t}$ samples and as well validated in the low BDT value region, which are shown in Fig. 7.8 and 7.9, respectively. The 30% uncertainty covers all possible discrepancies in the kinematical variables used in the analysis as well as in the distribution of the BDT value.

7.3.2 *Background caused by the mismeasurement of the lepton charge*

The process contributing to this category of background in the signal regions is primarily $t\bar{t}$ production, as observed in the simulation (115 events and 9 events for $t\bar{t}$ and DY MC, respectively). Nevertheless, to account for all possible processes that contribute in this background category, a partially data-driven technique is used for the estimation: opposite-charge dielectron and electron-muon events passing the full identification and kinematic selection in the analysis are reweighted with the probability of electron charge misidentification categorized in the electrons p_T and η. This probability is measured in a mix of $t\bar{t}$ and DY MC simulation.

To check the validity of this approach, the rate of charge mis-identification is first checked in dielectron events in $t\bar{t}$ and DY MC using information from a MC generator. The rate in this step is simply estimated as the the probability of the mismatch between the charge of the reconstructed electron and the generated electron. In the next step, the charge misidentification rate is measured in the DY MC simulation and data events containing dielectrons, the invariant mass of which lies 15 GeV close to the Z boson mass. In this case, the rate is computed as the number of same-charge events divided by the number of opposite-charge events. The charge misidentification rate is parametrized as a function of electron's η. Events with one central and one forward electron are further

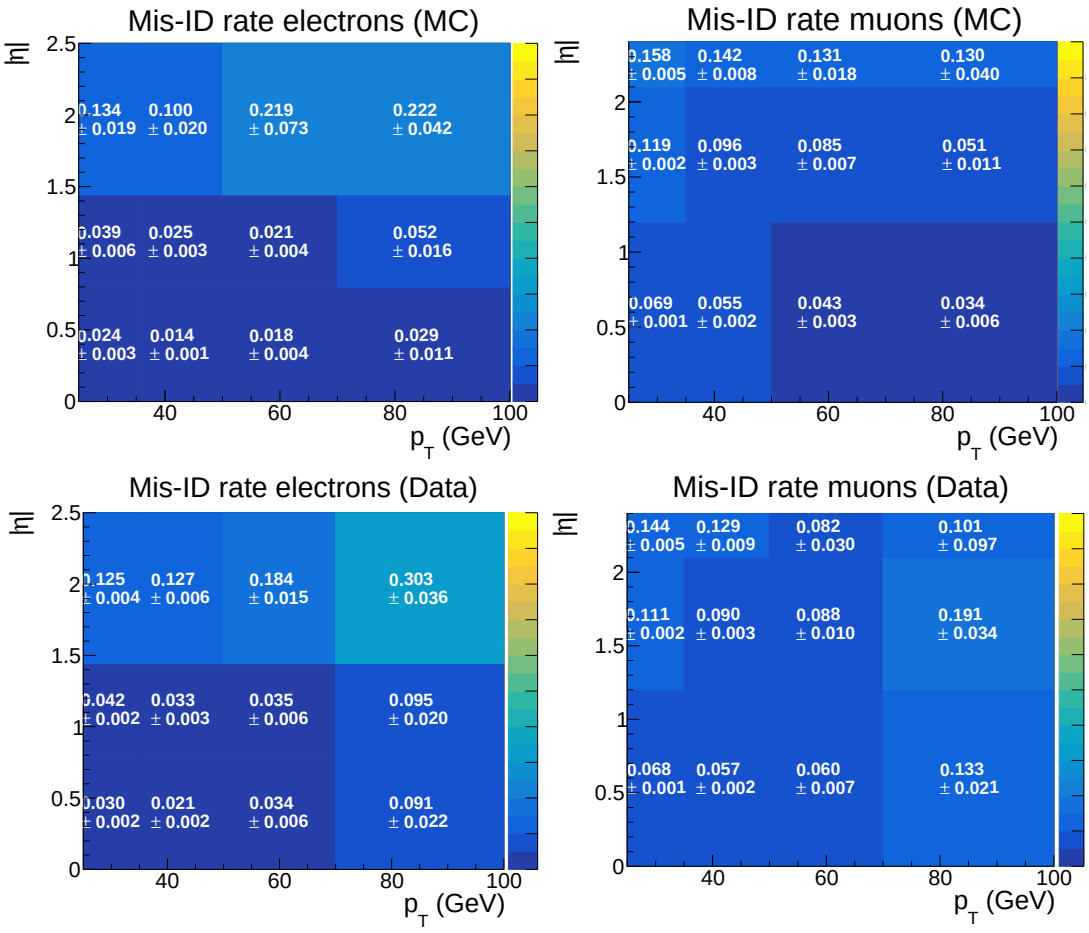


Figure 7.7: Misidentification rate measured for electrons (left) and muons (right) in QCD multijet events in MC (top) and data (right).

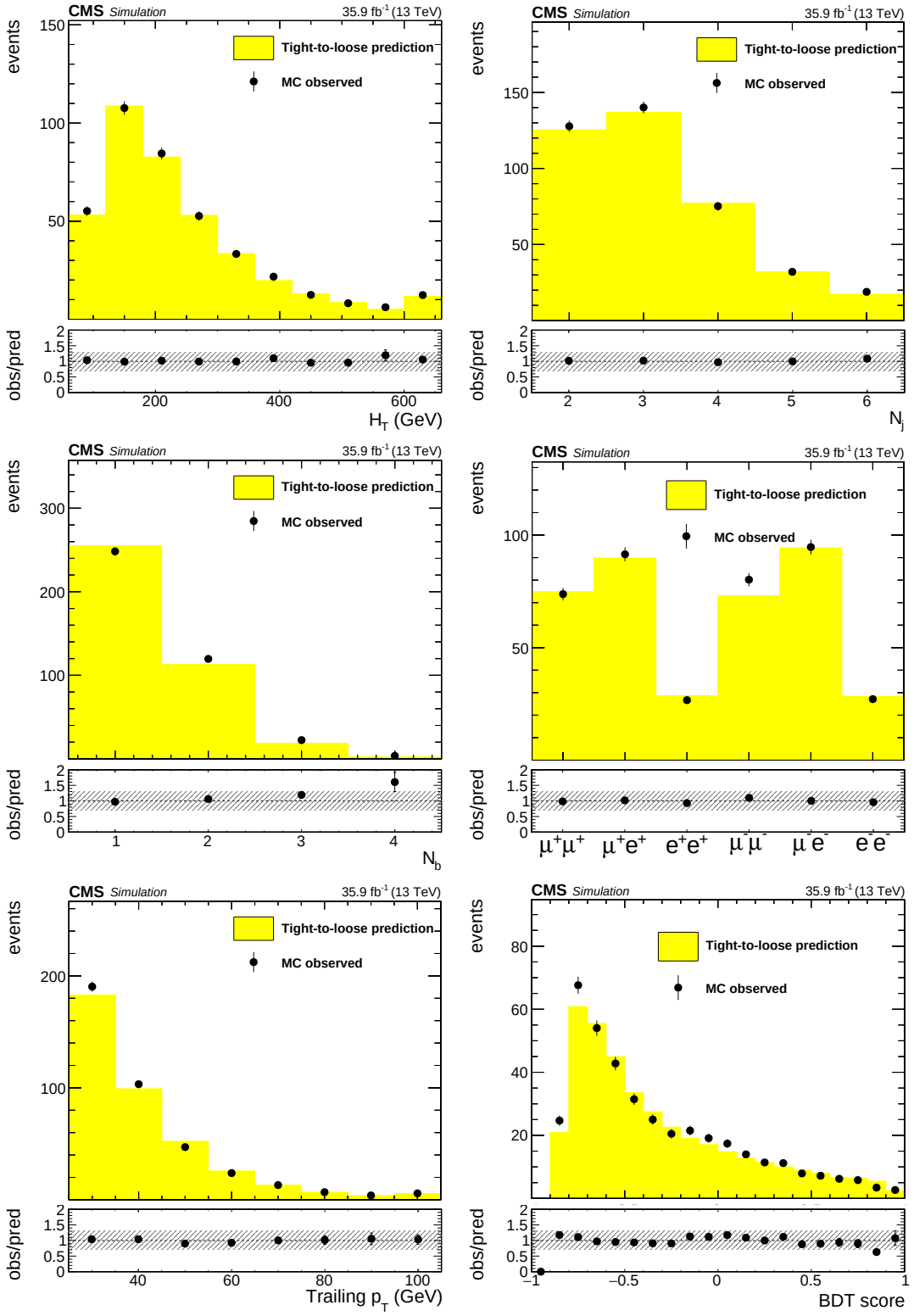


Figure 7.8: Validation for the nonprompt lepton background estimation technique in simulation. The distributions of H_{T}, N_{j}, N_{b}, the trailing lepton p_{T}, predicted number of events as a function of the flavor composition and BDT score of the event as obtained from a $\mathrm{t\bar{t}}$ sample are shown.

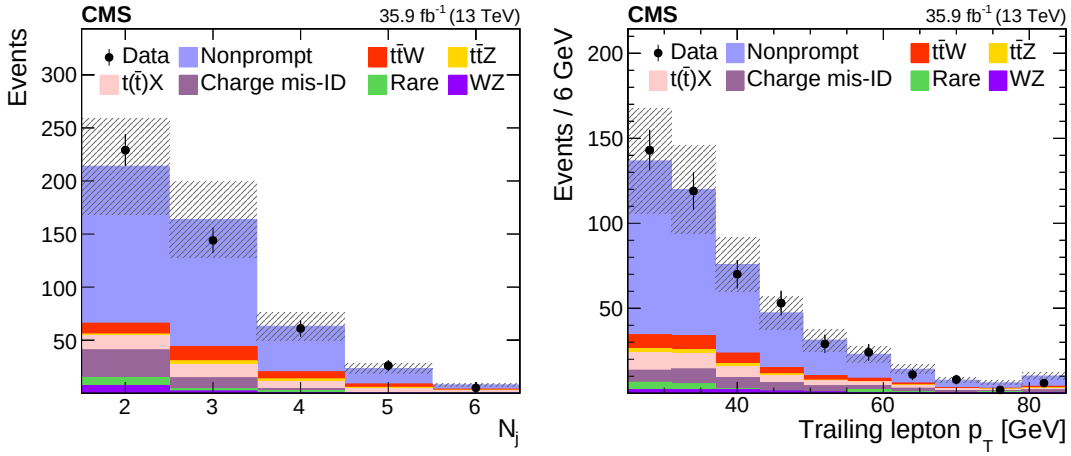


Figure 7.9: Predicted and observed yields as a function of N_j (left) and trailing lepton p_T (right) in BDT value less than 0 region enriched in nonprompt lepton background. The contribution from this type of background is estimated using the tight-to-loose technique described in Section 6.4.1. The shaded band represents the total uncertainty in the prediction of signal and background processes.

separated as to whether the leading or trailing electron is central. Table 7.2 shows measured charge misidentification rate in all described categories. From an agreement in the table, the 20% systematic uncertainty is assigned to this type of background.

Table 7.2: Charge misidentification rates for dielectron events. The rates are categorized by the η of the electrons, with the forward-central category containing a forward leading electron and the central-forward category containing a central leading electron. The first two columns were calculated using the generator-level information, while the third and fourth were calculated using events falling into the Z window. The uncertainty represents the statistical uncertainty of the measurement.

	$t\bar{t}$ (gen. info)	DY (gen. info)	DY (on-Z)	Data (on-Z)
Cent-Cent	0.00020 ± 0.00001	0.00014 ± 0.00002	0.00018 ± 0.00002	0.00024 ± 0.00001
Forw-Cent	0.00168 ± 0.00007	0.00151 ± 0.00009	0.00201 ± 0.00011	0.00214 ± 0.00006
Cent-Forw	0.00227 ± 0.00009	0.00187 ± 0.00010	0.00232 ± 0.00012	0.00223 ± 0.00006
Forw-Forw	0.00437 ± 0.00017	0.00425 ± 0.00019	0.00526 ± 0.00022	0.00510 ± 0.00013

7.3.3 *Background due to WZ production*

The WZ production is one of the background processes that can pass the event selection when one of the three leptons remains outside the detector acceptance or does not pass the selection described in Section 7.1. The contribution from this background is estimated using the MC simulation, while the normalization is verified in a dedicated three-lepton control region. It needs to be mentioned, that the $t\bar{t}W$ production was measured simultaneously with the $t\bar{t}Z$ cross section in the three-lepton and four lepton final states, where the lepton selection is slightly modified to maximize the sensitivity to the $t\bar{t}Z$ process. Therefore a common data control sample is formed using the lepton selection designed for the $t\bar{t}Z$ measurement. The full list of selection criteria is the

following: three leptons pass the identification and isolation criteria in the $t\bar{t}Z$ analysis; the p_T thresholds are 40, 20 and 10 GeV for the leading, sub-leading and trailing leptons; two leptons should form an OCSF pair, with an invariant mass of which is 10 GeV close to a Z boson; $N_b = 0$; $p_T^{miss} > 30$ GeV; and, finally, to reduce the contribution from nonprompt lepton background, the transverse mass calculated using the momentum vector of the lepton, that is not used in forming the Z boson candidate, and the p_T^{miss} is required to be a least 50 GeV.

After applying this selection the purity in WZ is around 80%. In this control region we measure a scale factor of WZ and take its uncertainty as a systematic uncertainty for the WZ normalization. The measured WZ scale factor is 0.94 ± 0.07, i.e. is compatible with 1. Based on this study we assign a rounded 10% systematic uncertainties on the normalization of the WZ production. An additional 20% uncertainty is concluded based on the small mismatch at high jet multiplicities ($N_j \geq 4$, see Fig. 7.10) between data and simulation. In addition to these two uncertainties, we assign theoretical and experimental uncertainties on the simulated acceptance for this process.

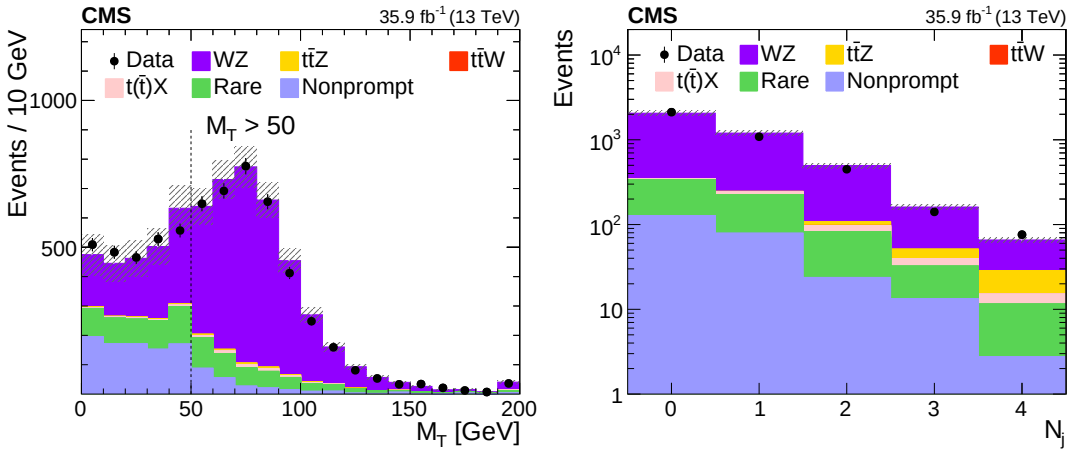


Figure 7.10: Predicted and observed yields as a function of M_T of third lepton (left) and N_j (right) in the WZ enriched control region. The contribution from the nonprompt lepton background is estimated using the tight-to-loose technique. The shaded band represents total uncertainty in the prediction of the background and the signal processes.

7.3.4 *Other processes*

Among the remaining backgrounds, the most significant processes in high N_j and N_b categories are the ones that produce at least one top quark in the final state, namely $t\bar{t}Z/\gamma^*$, $t\bar{t}H$ or tHq. These processes are simulated with various MC generators, see Table C.1. In the last category - rare SM backgrounds - the dominant contribution originates from the electroweak production of a W boson pair, that is modeled with the Madgraph generator. The contributions from all these processes are estimated by using MC simulations scaled to the integrated luminosity of 35.9 fb^{-1}. Appropriate scale factors are applied on the yields obtained with these simulations to take into account the differences between the efficiencies in simulation and data.

7.4 SYSTEMATIC UNCERTAINTIES

The procedure for estimation of systematic uncertainties is quite similar to the one described in Section 6.5. All uncertainties can be split into three components: experimental, theoretical and statistical. Table 7.3 lists all uncertainties relevant for the $t\bar{t}W$ analysis, gives the range for the effect on the yields of signal and background processes and describes impact on the measured $t\bar{t}W$ cross section. The procedure of obtaining the $t\bar{t}W$ cross section will be discussed later in the Section 7.5.

Table 7.3: Summary of the sources of uncertainties, their magnitude and the associated effect on the $t\bar{t}W$ cross section measurement. The second column indicates the magnitude of a given uncertainty and the third column indicates the corresponding effect on measured $t\bar{t}W$ production rate.

Source	Effect on the yield	Effect on $t\bar{t}W$ cross section measurement
Integrated luminosity	2.6%	4%
Jet energy scale and resolution	2 – 5%	3%
Trigger	2 – 4%	4.5%
b tagging heavy	1 – 3%	5%
b tagging light	1 – 3%	2%
PU reweighting	1%	1%
Lepton ID efficiency	3%	3%
Choice in μ_R and μ_F	1%	1%
PDF	1%	1%
Nonprompt lepton background	30%	4%
Charge misidentification	20%	3%
WZ cross section	10 – 20%	1%
$t\bar{t}X$ background	10%	4%
Rare SM background	50%	2%
Stat. unc. in nonprompt background	5 – 25%	4%
Total uncertainty		14%

This analysis makes use of the multivariate discriminant that complicates in some sense the estimation of the experimental uncertainties. The uncertainties on the JEC, scaled up and down for selected jets p_T and energy, are propagated not only to the N_j and N_b variables used in the event classification, but also to all other variables (H_T, M_T, p_T^{miss}, jets p_T) used for the multivariate discriminator. Due to the application of event selection, the first one has an acceptance effect while the second one causes the migration from one BDT-value category to another.

The corrections on the b tagging efficiencies for light and b flavor jets, which are parametrized as a function of p_T and η, are propagated to the N_b and therefore to the total BDT score. Therefore, similar to the JEC uncertainties, it has an effect both on the event acceptance and migration between the signal region bins due to the change in the BDT score.

Uncertainties on the lepton identification on average have an effect of 1.5% per lepton, even though they are parametrized as lepton p_T and η.

This dependency is propagated to the leptons p_T used in the multivariate discriminator and as well all variables dependent on the lepton p_T and η (M_T and

ΔR(trailing lepton, jet)). Nevertheless, the obtained effect of this type of uncertainty is rather flat in the signal region bins.

The procedure for estimating and propagating uncertainties related to the luminosity and PU reweighting doesn't change with respect to the one described in Chapter 6.5. All these experimental uncertainties are shown in Fig. 7.11 for events with a BDT score greater than 0. For simplicity, in these plots the yields from double positive and double negative charged categories are lumped together. The effect of uncertainties are shown for the signal process and the most important processes in the $t\bar{t}X$ category, namely $t\bar{t}Z$, $t\bar{t}H$ and tHq.

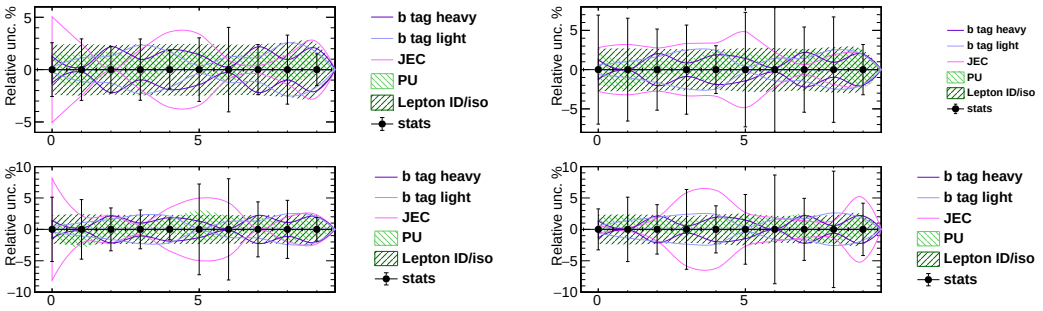


Figure 7.11: The value of the various components of the experimental systematic uncertainties in each signal region for events with a BDT score greater than 0 for the $t\bar{t}W$ (top left), $t\bar{t}Z$ (top right), $t\bar{t}H$ (bottom left) and tHq (bottom right) processes.

For triggers, besides the corrections explained in Section 7.1, an extra uncertainty of 2% is assigned to events with two electrons and electron-muon in the final state, while a 4% uncertainty is assigned to dimuon events. These uncertainties cover the statistical uncertainty of the measurement and as well the remaining mismatch between the measured efficiencies in data and simulation after applying corrections.

The theoretical uncertainties for processes in the $t\bar{t}X$ category are estimated at NLO QCD and electroweak accuracy and do not exceed 11% [129]. This uncertainty is taken as an uncertainty on the normalization for all processes in the $t\bar{t}X$ category. The procedure of obtaining the theoretical uncertainty that has an acceptance effect remains the same and gives an estimate for all backgrounds in $t\bar{t}X$ and rare SM processes categories. In general this effect doesn't exceed 1%. The total normalization uncertainty for the processes in the rare SM category is assigned to 50% due to the fact that most of the processes are simply have not been measured yet at the LHC.

One of the largest uncertainties originates from the imperfect knowledge of the nonprompt lepton background. The 30% is assigned to this type of background to cover the excesses in the closure test and control region in enriched nonprompt lepton background category. In addition to this, uncertainties on electroweak normalization and statistical precision of the nonprompt lepton background measurement are considered. The background related to the charge misidentification is a subject of 20% uncertainty obtained from the study explained in the Subsection 7.3.2.

7.5 RESULTS

Comparisons between data and predicted background in all signal regions are shown in Fig. 7.12 and 7.13. The same information is also presented in Tables 7.4 and 7.5 for the BDT score less and greater than 0 regions, respectively. These signal regions are used in the fit described in the subsequent subsection.

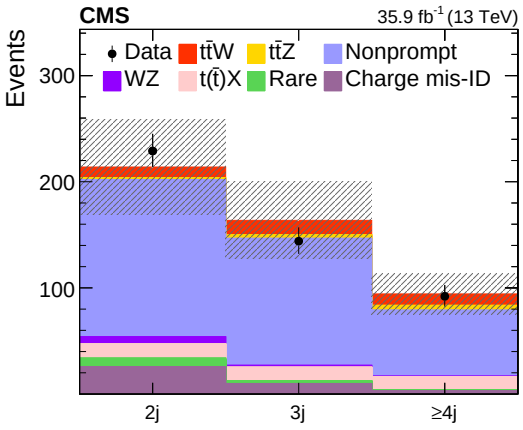


Figure 7.12: Predicted signal and background yields compared to observed data in the same-charged dilepton BDT score less than 0 categories. The hatched band shows the statistical and systematic uncertainties as explained in Section 7.4

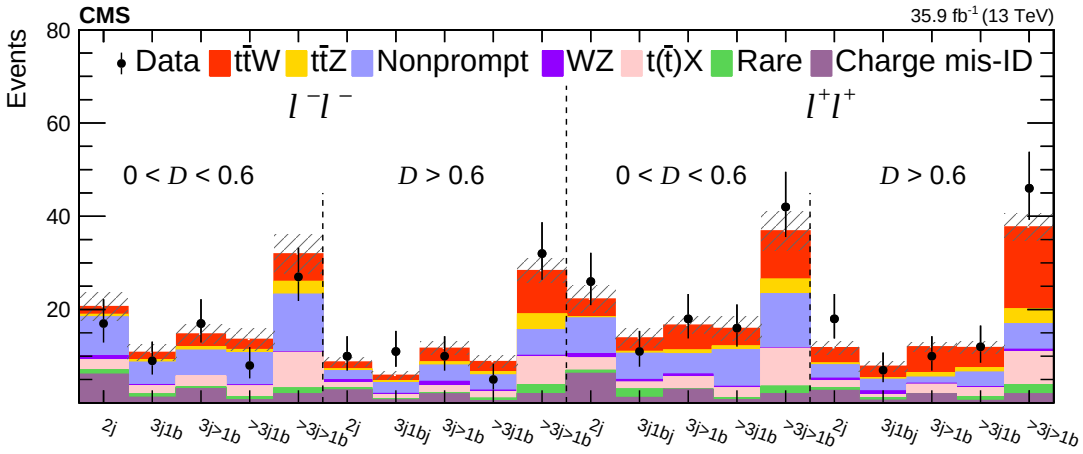


Figure 7.13: Predicted signal and background yields compared to observed data in the same-charged dilepton BDT score greater than 0 categories. The hatched band shows the statistical and systematic uncertainties as explained in Section 7.4

7.5.1 *Statistical procedure*

The statistical procedure used for this analysis and the analysis presented in the Chapter 8 follows the procedure described in [122]. Contrary to the SUSY analysis, the presence of signal is quantified by the background-only p-value. Such p-value can be interpreted as a probability of the background yields to fluctuate and give an excess of events as

Process	$N_j = 2$	$N_j = 3$	$N_j > 3$
Nonprompt	148.1 ± 44.6	119.6 ± 36.0	62.2 ± 18.8
$t\bar{t}Z$	1.7 ± 0.4	3.2 ± 0.8	4.6 ± 1.1
Total background	204.9 ± 45.0	152.5 ± 36.2	86.8 ± 18.9
$t\bar{t}W$	10.7 ± 0.3	14.3 ± 0.3	11.2 ± 0.3
Total	215.6 ± 45.0	166.8 ± 36.2	98.0 ± 18.9
Observed	229	144	92

Table 7.4: Predicted signal and background yields compared to observed data in the same-charge dilepton final state BDT score less than 0 categories. The uncertainty indicates the statistical and systematic uncertainties as explained in Section 7.4.

large or larger than the observed one. The test statistics used to quantify the excess is defined as a ratio of two likelihood functions:

$$q_0 = -2 \log \frac{\mathcal{L}(data|0, \hat{\theta}_0)}{\mathcal{L}(data|\hat{\mu}, \hat{\theta})}.$$

Here $\hat{\mu}$ is greater than 0 and is not constrained by any upper limit. Following a similar procedure as described in Subsection 6.6.1, the probability density function $f(q_0|0, \hat{\theta}^{obs})$ is used to generate pseudo-data for nuisance parameters around $\hat{\theta}^{obs}$ and event counts following Poisson probabilities under the assumption of the H_0 hypothesis. From such distribution, one can evaluate the p-value corresponding to a given experimental observation q_0^{obs} as follows:

$$p_0 = P(q_0 \geq q_0^{obs}|H_0) = \int_{q_0^{obs}}^{\infty} f(q_0|0, \hat{\theta}_0^{obs}) dq_0$$

The excess in observed data set can also be described with a significance, Z, that is linked to the p-value according to the following formula:

$$p = \int_Z^{\infty} \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}} dx$$

By convention, 5σ significance ($Z = 5$) is qualified as a discovery with a p-value equal to 2.87×10^{-7}. Considering the asymptotic behavior of the likelihood ratio test statistic q_0, the p-values and therefore corresponding significances can be obtained from the observed value q_0^{obs} itself, without a need for MC pseudo-data generation. In the asymptotic regime when all nuisance parameters can be profiled out of the likelihood, the distribution of q_0 is a mix of delta function at 0 and the χ^2-square function with one degree of freedom. The significances for $t\bar{t}W$ and $t\bar{t}Z$ are estimated with approximate formula [123] and to confirm the obtained number the significance for $t\bar{t}W$ was estimated by generating 22.5 millions MC pseudo-data points. The signal strength modifier described in Subsection 6.6.1 is extracted both for $t\bar{t}W$ and $t\bar{t}Z$ process by maximizing the likelihood function shown in Eq. 6.4.

		N_j	N_b	Background	$t\bar{t}W$	$t\bar{t}Z$	Total	Observed
$\ell^-\ell^-$	$0 < D < 0.6$	$= 2$	> 0	18.9 ± 3.0	1.7 ± 0.2	0.6 ± 0.2	20.6 ± 3.0	17
		$= 3$	$= 1$	9.2 ± 1.6	1.6 ± 0.2	0.5 ± 0.2	10.8 ± 1.6	9
			> 1	12.0 ± 1.9	2.8 ± 0.3	0.8 ± 0.2	14.8 ± 2.0	17
		> 3	$= 1$	11.4 ± 2.3	2.2 ± 0.2	0.7 ± 0.2	13.6 ± 2.3	8
			> 1	26.0 ± 4.1	5.9 ± 0.6	2.8 ± 0.7	31.9 ± 4.2	27
	$D > 0.6$	$= 2$	> 0	7.2 ± 1.0	1.6 ± 0.2	0.4 ± 0.1	8.7 ± 1.1	10
		$= 3$	$= 1$	4.6 ± 0.9	1.3 ± 0.2	0.3 ± 0.1	5.9 ± 0.9	11
			> 1	8.7 ± 1.4	3.0 ± 0.3	0.7 ± 0.2	11.7 ± 1.4	10
		> 3	$= 1$	6.6 ± 1.2	2.3 ± 0.3	0.7 ± 0.2	8.8 ± 1.2	5
			> 1	19.0 ± 2.3	9.3 ± 1.0	3.3 ± 0.8	28.3 ± 2.5	32
$\ell^+\ell^+$	$0 < D < 0.6$	$= 2$	> 0	18.4 ± 2.9	3.8 ± 0.4	0.3 ± 0.1	22.2 ± 2.9	26
		$= 3$	$= 1$	11.0 ± 2.0	2.9 ± 0.3	0.4 ± 0.1	13.9 ± 2.1	11
			> 1	11.3 ± 1.7	5.4 ± 0.6	0.8 ± 0.2	16.6 ± 1.8	18
		> 3	$= 1$	12.1 ± 2.6	3.8 ± 0.4	0.8 ± 0.2	15.9 ± 2.6	16
			> 1	26.5 ± 4.0	10.4 ± 1.1	3.1 ± 0.8	36.9 ± 4.1	42
	$D > 0.6$	$= 2$	> 0	8.4 ± 1.4	3.4 ± 0.4	0.4 ± 0.1	11.8 ± 1.4	18
		$= 3$	$= 1$	5.3 ± 1.0	2.5 ± 0.3	0.3 ± 0.1	7.8 ± 1.1	7
			> 1	6.4 ± 0.8	5.6 ± 0.6	1.0 ± 0.3	12.0 ± 1.0	10
		> 3	$= 1$	7.4 ± 1.2	4.4 ± 0.5	0.9 ± 0.3	11.8 ± 1.3	12
			> 1	20.1 ± 2.3	17.6 ± 1.8	3.2 ± 0.8	37.7 ± 2.9	46

Table 7.5: Predicted signal and background yields compared to observed data in the same-charge dilepton final state BDT score greater than 0 categories. The uncertainty indicates the statistical and systematic uncertainties as explained in Section 7.4.

7.5.2 *Measured ttW signal strength and the cross section*

The target of the analysis is to re-discover the $t\bar{t}W$ process and to measure its cross section. In order to extract the signal strength modifier μ with its associated 68% CL interval, a global maximum of the likelihood function described in Eq. 6.4 has been found. Furthermore, the likelihood function was maximized separately to find the best fit value for the $t\bar{t}W^+$ and $t\bar{t}W^-$ processes using only double-positive and double-negative events, respectively. The values of the best fit for the signal strength modifier that describes the data for $t\bar{t}W$, $t\bar{t}W^+$ and $t\bar{t}W^-$ are shown in Table 7.6.

$t\bar{t}W$	$1.23^{+0.19}_{-0.18}(\text{stat.})^{+0.20}_{-0.18}(\text{syst.})$
$t\bar{t}W^+$	$1.39^{+0.22}_{-0.22}(\text{stat.})^{+0.22}_{-0.22}(\text{syst.})$
$t\bar{t}W^-$	$0.91^{+0.34}_{-0.33}(\text{stat.})^{+0.29}_{-0.28}(\text{syst.})$

Table 7.6: Measured signal strength modifier for the $t\bar{t}W$, $t\bar{t}W^+$ and $t\bar{t}W^-$ processes.

The expected (observed) signal significance of 4.5 (5.3) standard deviations was found for the $t\bar{t}W$ processes in the 35.9 fb^{-1} data set collected in 2016. The expected (observed)

signal significances for $t\bar{t}W^+$ and $t\bar{t}W^-$ processes are calculated as well, being 4.2 (5.5) and 2.4 (2.3), respectively. At this point, we can claim the re-discovery of the $t\bar{t}W$ and $t\bar{t}W^+$ processes at $\sqrt{s} = 13$ TeV. The precision of the $t\bar{t}W$ production measurement has reached the level of 20%.

In order to extract the cross section of $t\bar{t}W$, $t\bar{t}W^+$ and $t\bar{t}W^-$ production, the cross section value used for the normalization of the signal sample is multiplied with the signal strength modifier. The $t\bar{t}W$ production is simulated with the MADGRAPH5_aMC@NLO (see Table C.1). The cross section estimated from the ME is equal to 0.2043 pb. In order to increase the statistics in the multilepton final state, only leptonic decays of a W boson are considered in the modeling, therefore the listed above cross section should be corrected by the branching ratio of $W \to \ell\nu$, which is equal to 0.326 [10]. The measured cross sections of the $t\bar{t}W$, $t\bar{t}W^+$ and $t\bar{t}W^-$ processes extracted from the same-charge dilepton final state are found to be:

$$\sigma(pp \to t\bar{t}W) = 0.77^{+0.12}_{-0.11}(\text{stat.})^{+0.13}_{-0.12}(\text{syst.})\text{pb}$$

$$\sigma(pp \to t\bar{t}W^+) = 0.58 \pm 0.09(\text{stat.})^{+0.09}_{-0.08}(\text{syst.})\text{pb}$$

$$\sigma(pp \to t\bar{t}W^-) = 0.19 \pm 0.07(\text{stat.}) \pm 0.06(\text{syst.})\text{pb}$$

The likelihood function described in Eq. 6.4 as a function of the cross section value along with the 68% and 95% CL intervals is shown in Fig. 7.14.

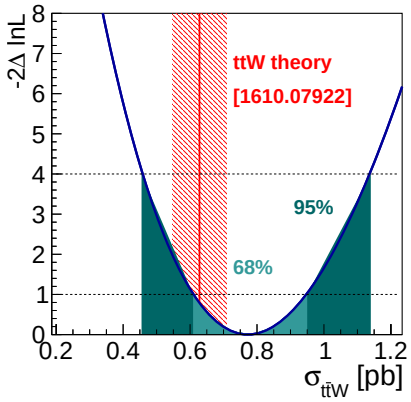


Figure 7.14: The measured cross section for the $t\bar{t}W$ process along with the 68% and 95% CL uncertainties and the theory prediction estimated in [129]

In Fig. 7.15 various kinematic distributions in the predicted and observed yields are presented in the $t\bar{t}W$ signal-enriched region, which is defined by requiring $N_j > 2$ and $N_b > 1$.

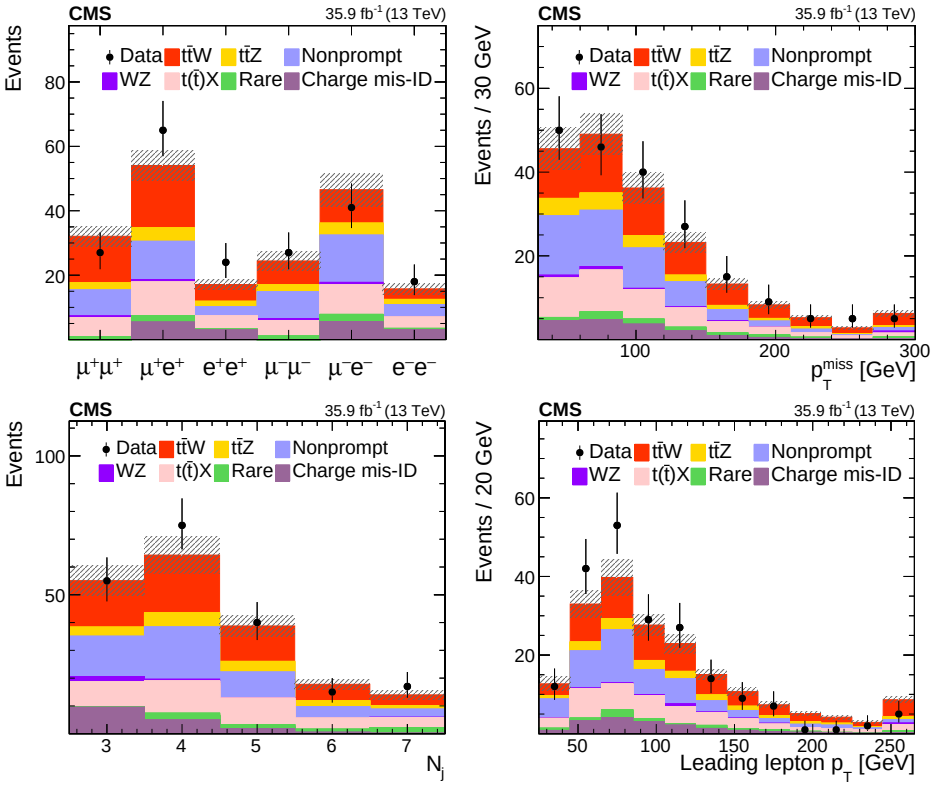


Figure 7.15: Predicted signal and background yields, as obtained from the fit, compared to observed data versus the flavor and the charge combination of leptons (upper left), p_{T}^{miss} (upper right), N_{j} (lower left), and the p_{T} of the leading lepton (lower right) in the the $\mathrm{t\bar{t}W}$ signal-enriched region with at least three jets and at least two b jets. The last bin in each distribution includes the overflow events, and the hatched band shows the total uncertainty associated with the signal and background predictions, as obtained from the fit.

8

MEASUREMENT OF THE TOP PAIR PRODUCTION IN ASSOCIATION WITH A Z BOSON

Using the collision data collected by the CMS detector during 2016, the inclusive $t\bar{t}Z$ cross section was measured simultaneously with the $t\bar{t}W$ process presented in Chapter 7. This measurement of the $t\bar{t}Z$ process was the subject of the PhD thesis worked out by Deniz Poyraz [130]. The measured cross section of the $t\bar{t}Z$ production is $0.99^{+0.09}_{-0.08}(\text{stat.})^{+0.12}_{-0.10}(\text{syst.})$pb and its significance is far beyond the discovery level. The measured uncertainty has reached 15% uncertainty level and is constrained by its systematic component, therefore to further improve the $t\bar{t}Z$ measurement we needed to concentrate our efforts on this component.

In this thesis, we report on an updated measurement of the $t\bar{t}Z$ cross section in three- and four-lepton channels with high jet and b jet multiplicity final states (see Fig. 8.1) using the proton-proton collision data collected by the CMS detector in 2016 and 2017 with a total integrated luminosity of 77.5 fb^{-1}. While the data analysis strategy remains similar to the one presented in [45], this new measurement benefits greatly from an improved lepton selection procedure based on MVA techniques and a more inclusive trigger selection.

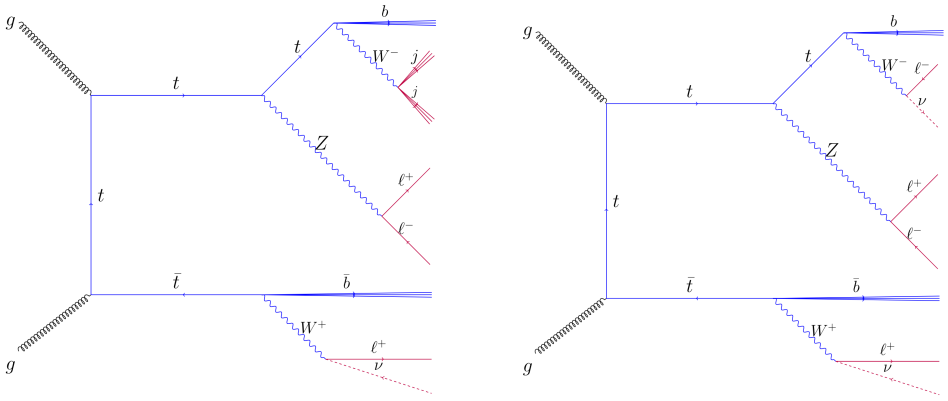


Figure 8.1: The $t\bar{t}Z$ decay process in the three-lepton (left) and four-lepton (right) final state. The leptons (denoted by ℓ) are accompanied by up to four jets (j), of which two are b-tagged jets(b), and up to two neutrinos (ν).

The obtained result challenges the theoretical uncertainty for the first time, where the theoretical uncertainties mainly originate from the choice of renormalization and factorization scales [129]. The inclusive cross section calculation also forms the basis for a subsequent differential cross section measurements. Both measurements are further utilized in setting stringent limits on NP within the EFT approach.

The results presented in this chapter are publicly available in a preliminary form called Physics Analysis Summary (PAS). The document is currently being prepared to be submitted to JHEP for publication in the coming months. The PAS is added to the thesis in the following after which we will give more details on the measurement done by the PhD candidate.

Available on the CERN CDS information server **CMS PAS TOP-18-009**

CMS Physics Analysis Summary

Contact: cms-pag-conveners-top@cern.ch 2019/03/09

Measurement of top quark pair production in association with a Z boson in proton-proton collisions at $\sqrt{s}$ = 13 TeV

The CMS Collaboration

Abstract

A measurement of the cross section of top quark pair production in association with a Z boson using proton-proton collisions at a center-of-mass energy of 13 TeV at the LHC is performed. The data sample corresponds to an integrated luminosity of 77.5 fb^{-1}, collected by the CMS experiment during 2016 and 2017. The measurement is performed in the three- and four-lepton final states. The production cross section is measured to be $\sigma(\mathrm{t\bar{t}Z}) = 1.00\,^{+0.06}_{-0.05}\,(\mathrm{stat})\,^{+0.07}_{-0.06}\,(\mathrm{syst})$ pb. Differential cross sections are measured as a function of the transverse momentum of the Z boson and the angular distribution of the decay lepton. New stringent limits on the anomalous couplings of the top quark to the Z boson are obtained, including estimates of Wilson coefficients of standard model effective field theory.

1 Introduction

The large amount of proton-proton (pp) collision data at a center-of-mass energy of 13 TeV at the LHC allows precision measurements of standard model (SM) processes with very small production rate. Production of top quark-antiquark pairs ($t\bar{t}$) in association with a Z ($t\bar{t}Z$) boson is among such processes. The theoretical cross sections at next-to-leading order (NLO) in quantum chromodynamics (QCD) for $t\bar{t}Z$ production at $\sqrt{s} = 13\,\text{TeV}$ is $0.839 \pm 0.101\,\text{pb}$ [1]. Precise measurements of the inclusive and differential cross sections of the $t\bar{t}Z$ process are of particular interest because these topologies can receive sizable contributions from physics beyond the standard model (BSM) [2, 3]. The $t\bar{t}Z$ production is the most sensitive process for directly measuring the coupling of the top quark to the Z boson. Furthermore, this process is an important background to several searches for BSM physics, as well as to measurements of SM processes, such as $t\bar{t}$ production in association with the Higgs boson ($t\bar{t}H$).

The inclusive cross section for $t\bar{t}Z$ production was measured by both the CMS and ATLAS collaborations using collision data at $\sqrt{s} = 13\,\text{TeV}$ collected in 2016 corresponding to an integrated luminosity of about $36\,\text{fb}^{-1}$. The CMS Collaboration used collision events containing three or four charged leptons (electrons or muons) and reported a measured value $\sigma(t\bar{t}Z) = 0.99\,^{+0.09}_{-0.08}\,(\text{stat})\,^{+0.12}_{-0.10}\,(\text{syst})\,\text{pb}$ [4], while the ATLAS Collaboration, in addition to the three- and four-lepton channels, included events with two leptons and presented a measurement with a similar precision [5].

In this note, we report on an updated measurement of the $t\bar{t}Z$ cross section in three- and four-lepton final states using the pp collision data collected by the CMS detector in 2016 and 2017 with a total integrated luminosity of $77.5\,\text{fb}^{-1}$. While the data analysis strategy remains similar to the one presented in [4], this new measurement benefits largely from an improved lepton selection procedure based on multivariate analysis techniques and a more inclusive trigger selection. In addition to the inclusive cross section, the differential cross section is measured as a function of the transverse momentum of the Z boson $p_{\text{T}}(\text{Z})$ and $\cos\theta^*_{\text{Z}}$. The latter observable is the cosine of the angle between the Z boson and the negatively charged lepton from the Z boson leptonic decay as measured in the Z rest frame, and is sensitive to its polarisation.

Because of the key role of the top quark interaction with the Z boson in many BSM models [6–12], the differential cross section measurements can be used to constrain anomalous $t\bar{t}Z$ couplings. We pursue two different interpretations to this end. First, we use a Lagrangian containing anomalous couplings [13] to obtain bounds on the vector and axial-vector currents, as well as the electroweak dipole moments of the top quark. The interpretation is extended in the context of top quark effective field theory (EFT) [14], and we constrain the Wilson coefficients of the relevant BSM operators of mass dimension 6. There are 59 operators, among which we identify the four most relevant linear combinations, as described in Ref. [15].

2 Data samples and object selection

The data sample used in this measurement corresponds to a total integrated luminosity of $77.5\,\text{fb}^{-1}$ of pp collision events collected by the CMS detector during 2016 and 2017. To incorporate the LHC running conditions, as well as the CMS detector performance in each year properly, the two data sets were analyzed independently with appropriate calibrations applied and combined at the final stage to extract the cross section value. The combined result is then used in the unfolding procedure.

Simulated Monte Carlo (MC) events are used to model the signal selection efficiency, to test

2

the background prediction techniques, and to predict some of the background yields. Two sets of simulated events for each process are used in order to match the different data-taking conditions in 2016 and 2017. Events for the $\mathrm{t\bar{t}Z}$ signal process and a variety of background processes, including production of Z+jets and triple vector boson (VVV) events, are simulated at next-to-leading order (NLO) in QCD using the MADGRAPH5_aMC@NLO generator [16]. The NLO POWHEG v2 [17] generator is used for the production of the $\mathrm{t\bar{t}}$ and WW processes, as well as for processes involving the Higgs bosons (H), produced in vector boson fusion (VBF), or in association with vector bosons or top quarks. The NNPDF3.0 (NNPDF3.1) [18, 19] parton distribution functions (PDFs) are used for the hard process for simulating the 2016 (2017) samples. Table 1 gives an overview of the event generators, PDF sets, and cross section calculations that are used for the signal and background processes. For all processes, the parton showering and hadronization are simulated using PYTHIA 8.2 [20, 21]. The modeling of the underlying event is generated using the CUETP8M1 [22, 23] and CP5 tunes **[GEN-17-001, currently in CWR]** for simulated samples corresponding to the 2016 and 2017 data sets, respectively. In 2016, the CUETP8M2 and CUETP8M2T4 tunes [24] are used for the $\mathrm{t\bar{t}H}$ and $\mathrm{t\bar{t}VV}$ samples. Double counting of the partons generated with MADGRAPH5_aMC@NLO and those with PYTHIA is removed using the MLM [25] and the FXFX [26] matching schemes in leading order (LO) and NLO generated events, respectively.

The theoretical $\mathrm{t\bar{t}Z}$ cross section used to normalize the expected signal yield, as well as to rescale the measured signal strength to obtain the final cross section, is computed for $\sqrt{s} = 13$ TeV at NLO in QCD using MADGRAPH5_aMC@NLO and the PDF4LHC15 recommendation [27]. It is found to be 0.839 ± 0.101 pb [1]. In this calculation, the phase space is limited to invariant masses greater than 10 GeV for all pairs of leptons.

Table 1: Event generators used to simulate events for the various processes. The version of the NNPDF set used for the hard process is shown for samples corresponding to the 2016 (2017) data sets.

Process	Cross section calculation	Event generator	Perturbative order	NNPDF version
$\mathrm{t\bar{t}Z}$, tZq, $\mathrm{t\bar{t}W}$, WZ, $\mathrm{t\bar{t}}\gamma^{(*)}$, $\mathrm{W}\gamma^{(*)}$, $\mathrm{Z}\gamma^{(*)}$, Z+jets, VVV	NLO	MADGRAPH5_aMC@NLO	NLO	3.0NLO (3.1NNLO)
$\mathrm{gg} \to \mathrm{ZZ}$	NLO [28]	MCFM v7.0.1 [29], JHUGEN [30]	LO	3.0LO (3.1LO)
$\mathrm{q\bar{q}} \to \mathrm{ZZ}$	NNLO [31]	POWHEG v2 [32, 33]	NLO	3.0NLO (3.1NNLO)
WH, ZH	NLO	POWHEG v2 MINLO HVJ [34], JHUGEN [30]	NLO	3.0NLO (3.1NNLO)
WW, VBF H	NLO	POWHEG v2	NLO	3.0NLO (3.1NNLO)
$\mathrm{t\bar{t}H}$	NLO	POWHEG v2 [35]	NLO	3.0NLO (3.1NNLO)
$\mathrm{t\bar{t}}$	NNLO+NNLL [36]	POWHEG v2	NLO	3.0NLO (3.1NNLO)
$\mathrm{t\bar{t}VV}$, tHW, tHq, tWZ	LO	MADGRAPH5_aMC@NLO	LO	3.0LO (3.1NNLO)

All events are processed through a simulation of the CMS detector based on GEANT4 [37] and are reconstructed with the same algorithms as used for data. Simultaneous pp collisions in the same or nearby bunch crossings, referred to as pileup (PU), are also simulated. The distribution used in the simulation is chosen to match the one observed in the data. The average PU in the years 2016 and 2017 was 27 and 38, respectively.

The particle-flow (PF) algorithm [38] is used for particle reconstruction and identification. Jets are reconstructed by clustering PF candidates using the anti-k_{T} algorithm [39] with a distance parameter $R = 0.4$. The influence of PU is mitigated through a charged hadron subtraction technique, which removes the energy of charged hadrons not originating from the primary pp interaction vertex (PV) [40]. The PV is defined as the vertex for which the sum of the physics objects p_{T}^2 has the largest value. Jets are calibrated in simulation, and separately in data, accounting for energy deposits of neutral particles from PU and any nonlinear detector response.

Jets with $p_T > 30\,\text{GeV}$ and $|\eta| < 2.4$ are selected for the analysis. Jets are identified as originating from b quarks using the Deep CSV algorithm [41]. This algorithm achieves a p_T-averaged efficiency of 70% for b quark jets to be correctly identified, while retaining a mistake rate of 12% for charm quark jets and 1% for jets originating from u, d, or s quarks or from gluons. The observable p_T^{miss} is computed as the magnitude of the negative vector sum of the p_T of all PF candidates.

Lepton identification and selection is a critical ingredient in this measurement. To achieve an effective rejection of the nonprompt leptons while keeping high efficiency for prompt leptons a multivariate analysis has been developed separately for electrons and muons similar to the one presented in Ref. [42]. Prompt leptons are those originating from direct W or Z boson decays, while nonprompt are those that are either misidentified jets or genuine leptons resulting from semileptonic decays of hadrons with heavy-flavor quarks. A boosted decision tree (BDT) classifier is used via the TMVA toolkit [43] for multivariate analysis. In addition to the lepton p_T and $|\eta|$, the training uses several discriminating variables, e.g., the kinematic properties of the closest jet to the lepton; the impact parameter in the transverse plane of the lepton tracks with respect to the PV; geometric matching of the track in the silicon tracker with the signals measured in the muon chambers; the ECAL shower shape for electrons; two variants of relative isolation – one computed with a fixed ($R = 0.3$) and the other with a variable cone size depending on the lepton p_T. The relative isolation is defined as the scalar sum p_T of the particles within a cone around the lepton direction, divided by the lepton p_T. Placing a stringent requirement on the BDT output, an increase of up to 15% in prompt lepton selection efficiency with respect to the one presented in [4] is achieved, while reducing the nonprompt lepton selection efficiency by about a factor 2 to 4, depending on the lepton p_T. Electrons (muons) passing the BDT selection and having $p_T > 10\,\text{GeV}$ and $|\eta| < 2.5(2.4)$ are selected. In order to avoid double counting, jets within a cone of $\Delta R < 0.4$ around the selected leptons are discarded.

3 Event selection and observables

Events are selected using a suite of online triggers each of which requires the presence of 1, 2 or 3 leptons. For events selected by the triggers that require at least one electron or muon, the transverse momentum threshold for electrons (muons) was 27 (24) GeV during 2016 and 32 (27) GeV in 2017. For triggers that require the presence of at least two leptons, the p_T thresholds are 23 (17) GeV for the highest p_T (leading) and 12 (8) GeV for the second-highest p_T (subleading) electron (muon). This strategy ensures an overall trigger efficiency higher than 98% over the entire 2016 and 2017 data sets. These efficiencies are measured in data samples with an independent trigger selection and compared to those obtained in simulation. The measured differences are mitigated by reweighting the simulation by appropriate factors that remain smaller than 2% (3%) in the 2016 (2017) data set.

Events with exactly three leptons ($\mu\mu\mu$, $\mu\mu$e, μee, or eee) satisfying $p_T > 10, 20, 40\,\text{GeV}$ or exactly four leptons ($\mu\mu\mu\mu$, $\mu\mu\mu$e, $\mu\mu$ee, μeee, or eeee) with $p_T > 10, 10, 10, 40\,\text{GeV}$ are analyzed separately. In both categories, exactly one oppositely charged and same-flavour lepton pair with an invariant mass consistent with the Z boson hypothesis is required, namely, for three (four)-lepton category $|m(\ell\ell) - M(Z)| < 10\ (20)\,\text{GeV}$. This selection reduces the backgrounds from events having zero or more than one Z boson. Events containing zero jets are rejected. The measurement uses the jet multiplicity, N_j, distributions in different event categories depending on the number of b-tagged jets, N_b, in the event. For the three-lepton channel these are $N_b = 0, 1, \geq 2$, while for the four-lepton channel these categories are limited to $N_b = 0, \geq 1$. In addition to the signal enriched regions, the analysis makes use of several data control re-

gions to validate the background predictions, as well as to control the systematic uncertainties associated with them. The details are given in Section 4.

Two kinematic observables, the transverse momentum $p_{\mathrm{T}}(\mathrm{Z})$ of the Z boson, and $\cos\theta^*_{\mathrm{Z}}$, the angle between the negatively charged lepton and the Z candidate measured in the Z rest frame, are used to constrain anomalous t-Z interactions. In addition, the $\mathrm{t\bar{t}Z}$ differential cross section is measured as a function of these observables.

4 Background predictions

Several SM processes contribute to the three- and four-lepton final states. The $\mathrm{t\bar{t}Z}$ process typically produces events with larger jet and b jet multiplicities. In contrast, events with $N_{\mathrm{b}} = 0$ are predominantly due to background processes. Following closely the methodologies used in Ref. [4], the separation between signal and backgrounds is obtained from a nuisance parameter fit. In the fit, the contributions from the various background processes are allowed to float within their uncertainties.

The main contributions to the total background arise from processes with at least one top quark produced in association with a W, Z, or a H boson, i.e., $\mathrm{t\bar{t}H}$, $\mathrm{t\bar{t}W}$, tWZ, tqZ, tHq, tHW, $\mathrm{t\bar{t}VV}$, and $\mathrm{t\bar{t}t\bar{t}}$. They are collectively denoted as $\mathrm{t}(\bar{\mathrm{t}})\mathrm{X}$. We consider both the theoretical and experimental systematic uncertainties in the background yields of the $\mathrm{t}(\bar{\mathrm{t}})\mathrm{X}$ category. The theoretical uncertainties in the inclusive cross section are around 11% [16, 44, 45]. This uncertainty is evaluated by varying the renormalization and factorization scales (μ_{R} and μ_{F}) for the matrix element and parton shower by a factor of two up and down, as well as the uncertainties due to the choice of the PDF. Among these processes, the tqZ cross section was recently measured by the CMS Collaboration with a total precision of 15% [46]. Thus we use this measurement and its uncertaintiy for the tqZ cross section.

The WZ production constitutes the second-largest background contribution, in particular for events with three leptons, while in the four-lepton category, ZZ production becomes substantial. For both these processes, the overall production rate, as well as several relevant kinematic distributions predicted from simulation, can be tested in respective data samples that are orthogonal to the signal selection. Events with three leptons, two of which form a same-flavor pair with opposite charge and satisfy $|m(\ell\ell) - M(\mathrm{Z})| < 10\,\mathrm{GeV}$, and $N_{\mathrm{b}} = 0$ are used to validate the WZ background prediction. Four-lepton events with two Z boson candidates are used to constrain the uncertainties in the prediction of the ZZ yield. Figure 1 presents the observed and predicted event yields in these regions. Good agreement is observed. An overall normalization uncertainty of 10% is assigned [47], while for events with $N_{\mathrm{j}} \geq 3$ an additional 20% uncertainty is assigned to the WZ background prediction. We also estimate potential mismodeling effects associated with the WZ production with heavy-flavor quark pairs originating from gluon splitting in a data sample enriched with Z+jets events. In a control region with a Z boson candidate and two b-tagged jets, a comparison of the distributions of the angle between the two b jets in data and simulation is used to assess the uncertainty due to possible mismodeling. It is estimated to be less than 20%. Propagating this uncertainty to the rate of simulated WZ events with gluon splitting, the uncertainty in the rate of WZ events with $N_{\mathrm{b}} \geq 1$ is estimated to be 8%.

Backgrounds with leptons from heavy-flavor hadron decays, muons from long-lived light meson decays, or electrons from unidentified photon conversions are estimated from data. Such backgrounds, referred to as "nonprompt" in the following, mainly originate from $\mathrm{t\bar{t}}$ or $\mathrm{Z} \to \ell\ell$ events, in which an additional nonprompt lepton arises from a semileptonic decay of a b

hadron. The lepton selection employed in this analysis specifically targets the reduction of nonprompt lepton backgrounds to a subdominant level, while keeping the signal efficiency high. The lepton selection is validated in simulation and with data control samples. Events with three leptons without a Z boson candidate provide a suitable data control sample. Figure 2 shows the predicted and observed yields in this control region for different lepton flavors as a function of the p_T of the lowest-p_T lepton and N_b, where good agreement between predicted and observed yields is obtained. Based on these studies, a systematic uncertainty of 30% is assigned, while the statistical uncertainty in the prediction of this background ranges between 5–50%, depending on the measurement bin.

A small contribution to the total background comes from the WW, WWW, WWZ, WZZ, and ZZZ processes. We group them in the "rare" category as these have relatively small production rates. Processes that involve a photon ($Z\gamma$, $t\gamma$, and $t\bar{t}\gamma^{(*)}$) are denoted by $X\gamma$. The contribution from both of these categories to the selected event count is evaluated using simulated samples described in Section 2. As in the case of the $t(\bar{t})X$ backgrounds, scale factors are applied to account for small differences between simulation and data due to trigger selection, lepton identification, jet energy corrections, and b jet selection efficiency. The overall uncertainty in the normalization of the "rare" background category is estimated to be 50% [1, 48], while for $X\gamma$ it is 20% [49]. The statistical uncertainty from the finite size of the simulated background samples are typically small and are typically around 5%, reaching 100% only in some extreme jet multiplicity regions.

5 Systematic uncertainties

The systematic uncertainties affecting the signal selection efficiency and background yields are summarized in Table 2. The table indicates both the range of variations observed in the different bins of the analysis and their impact on the final result. The table also indicates whether the uncertainties are treated as fully correlated between the two data sets, 2016 and 2017.

The uncertainty in the integrated luminosity measurement in 2016 and 2017 data is 2.5% [50]. Simulated events are reweighted according to the distribution of the expected number of interactions at each bunch crossing corresponding to a total inelastic pp cross section of 69.2 mb. The uncertainty in the latter, which affects the PU estimate, is 5% [51] and leads to a 2% uncertainty in the expected yields.

The trigger selection efficiencies in the simulation are found to be in good agreement with the ones measured in data. Based on this comparison, a 2% uncertainty is assigned to the yields obtained in the simulation. Lepton selection efficiencies are measured using a "tag-and-probe" method [52, 53] in bins of lepton p_T and η, and are found to be higher than 60 (95)% in events with three leptons, and 80 (95)% with four leptons for $p_T < 25$ (> 25) GeV. These measurements are performed separately in data and simulation. The differences between these two measurements, which are around 1% and reach 10% for leptons with $p_T < 20$ GeV, are used to scale the yields obtained in the simulation. The systematic uncertainties related to this source vary between 4.5–6% in the signal and background yields.

Uncertainties in the jet energy calibration are estimated by shifting the jet energy in the simulation up and down by one standard deviation. Depending on p_T and η, the uncertainty in jet energy scale changes by 2–5% [54]. For the signal and backgrounds modeled through simulation, the uncertainty in the measurement is determined from the observed differences in yields with and without the shift in jet energy. The same technique is used to calculate the uncertainties due to jet energy resolution, which are found to be less than 1%. The b tagging efficiency

6

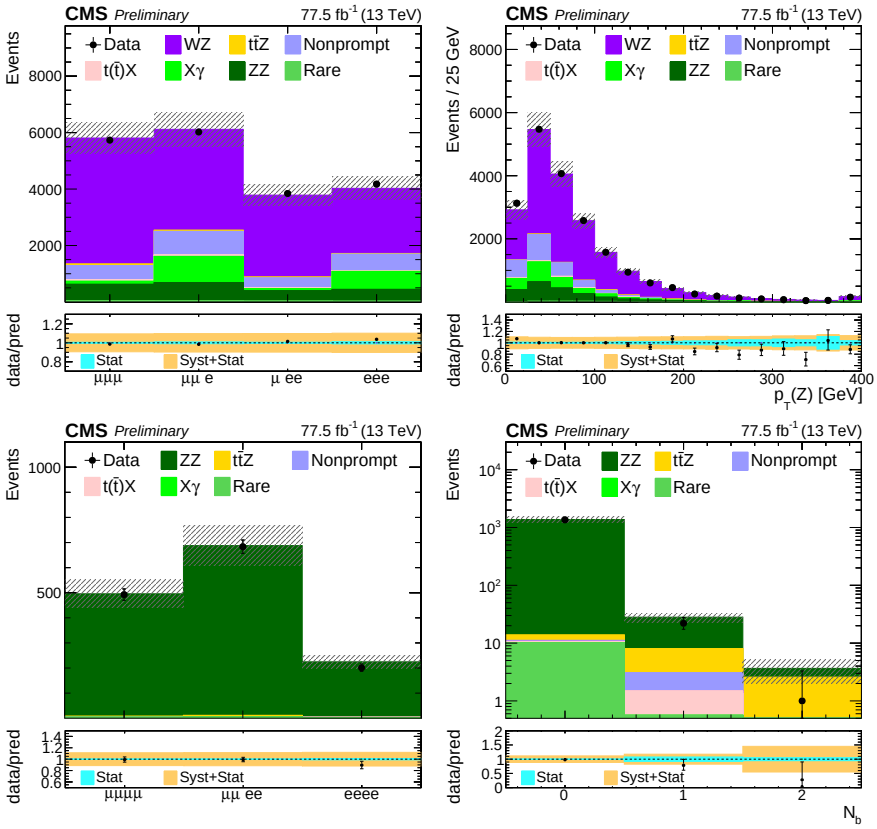


Figure 1: Distributions of the predicted and observed yields versus lepton flavor (upper left), and the reconstructed transverse momentum of the Z boson candidates (upper right) in the WZ-enriched data control region, and versus lepton flavor (lower left) and number of b jets (lower right) in the ZZ-enriched control region. The shaded band represents the total uncertainty in the prediction of the signal and background. The lower panels show the ratio of the data to the theoretical predictions. The inner band gives the statistical uncertainty in the ratio, and the outer band the total uncertainty.

in the simulation is corrected using scale factors determined from data [55, 56]. These are estimated separately for correct and misidentified tagging rates, and each results in an uncertainty about 1–4%, depending on the b tag multiplicity.

To estimate the theoretical uncertainties due to the choice of μ_R and μ_F, each of these parameters is varied independently up and down by a factor of 2, ignoring the anticorrelated variations. The envelope of the acceptance variations is taken as the systematic uncertainty in each search bin and is found to be smaller than 4%. The different variations in the NNPDF30 PDF set [18] are used to estimate the corresponding uncertainty in acceptance for the differential cross section measurement, which is typically less than 1%. The uncertainty associated with the choice of PDF for the anomalous coupling and EFT interpretations is estimated by using several PDFs and assigning the maximum differences as the quoted uncertainty, following the PDF4LHC prescription with the MSTW2008 68% CL NNLO, CT10 NNLO, and NNPDF2.3 5f FFN PDF sets (see Ref. [27] and references therein, as well as Refs. [57–59]). In the parton shower simulation, the uncertainty from the choice of μ_F is estimated by varying the scale of initial- and final-state radiation up by factors of 2 and $\sqrt{2}$ and down by factors of 0.5 and $1/\sqrt{2}$,

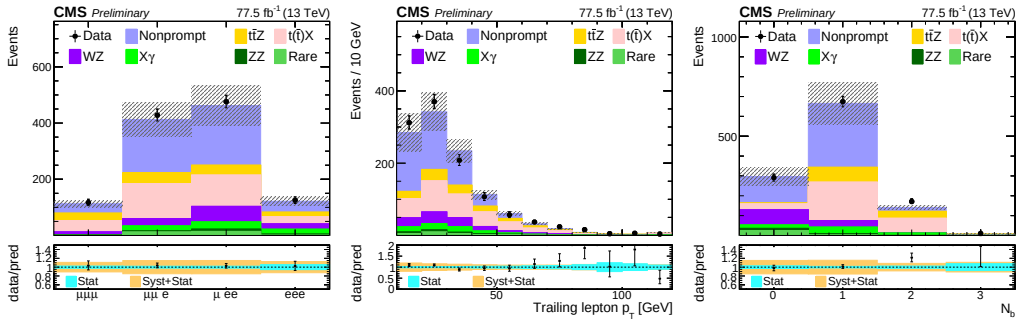


Figure 2: Distributions of the predicted and observed yields in regions enriched with nonprompt lepton backgrounds in $\mathrm{t\bar{t}}$-like processes as a function of the lepton flavors (left), the p_{T} of the lowest-p_{T} lepton (middle), and N_{b} (right). The shaded band represents the total uncertainty in the prediction of the background and the signal processes. The lower panels show the ratio of the data to the predictions from simulation. The inner band gives the statistical uncertainty in the ratio, and the outer band the total uncertainty. See Section 4 for the definition of each background category.

respectively, as suggested in Ref. [22]. The default setup in PYTHIA includes a multiple-parton-interaction-based (MPI) model of color reconnection with early resonance decays switched off. To estimate the uncertainty from this choice of model, the analysis is repeated with three other color reconnection models within PYTHIA: the MPI-based scheme with early resonance decays switched on, a gluon-move scheme [60], and a QCD-inspired scheme [61]. The total uncertainty from color reconnection modeling is estimated by taking the maximum deviation from the nominal result and amounts to 1.5%.

6 Results

6.1 Inclusive cross section measurement

The signal extraction is performed using various jet and b jet categories from events with three and four leptons. The observed data, as well as the predicted signal and background yields are shown in Fig. 3. The statistical procedure to extract the cross section is detailed in Refs. [62–65]. The observed yields and background estimates in each analysis category, and the systematic uncertainties described in Section 5 are used to construct a binned likelihood function $L(r,\theta)$ as a product of Poisson probabilities of all bins. The parameter r is the signal-strength modifier and θ represents the full suite of nuisance parameters. The test statistic is the profile likelihood ratio, $q(r) = -2\log L(r,\hat{\theta}_{\mathrm{r}})/L(\hat{r},\hat{\theta})$, and an asymptotic approximation is used to extract the fitted cross section, the associated uncertainties, and the significance of the observation of the signal process [62–65], where $\hat{\theta}_{\mathrm{r}}$ reflects the values of the nuisance parameters that maximize the likelihood function for signal strength r. The quantities $\hat{r}$ and $\hat{\theta}$ are the values that simultaneously maximize L. The measured combined signal strength is $1.19\,^{+0.07}_{-0.06}\,(\mathrm{stat})\,^{+0.08}_{-0.07}\,(\mathrm{syst})\,^{+0.11}_{-0.12}\,(\mathrm{theo})$, while the signal strengths separately for the three- and four-lepton channels are shown in Table 3. The measured $\mathrm{t\bar{t}Z}$ cross section is defined in the full phase space in which $m(\ell\ell) > 10$ GeV. It is measured to be

$$\sigma(\mathrm{pp} \to \mathrm{t\bar{t}Z}) = 1.00\,^{+0.06}_{-0.05}\,(\mathrm{stat})\,^{+0.07}_{-0.06}\,(\mathrm{syst})\,\mathrm{pb}.$$

The individual contributions to the total systematic uncertainty in the measured cross section

8

Table 2: Summary of the sources, magnitudes, treatments, and effects of the systematic uncertainties in the final $\mathrm{t\bar{t}Z}$ cross section measurement. The first column indicates the source of the uncertainties, the second column shows the corresponding input uncertainty range on each background source and the signal. The third column indicates how correlations are treated between uncertainties in the 2016 and the 2017 data, and the fourth column shows the resulting uncertainty in the $\mathrm{t\bar{t}Z}$ cross section.

Source	Uncertainty range (%)	Correlated in 2016 and 2017	Impact on the $\mathrm{t\bar{t}Z}$ cross section (%)
Integrated luminosity	2.5	×	2
PU modeling	1–2	✓	1
Trigger	2	×	2
Lepton ID efficiency	4.5–6	✓	4
Jet energy scale	1–9	✓	2
Jet energy resolution	0–1	✓	1
B tagging light flavor	0–4	×	1
B tagging heavy flavor	1–4	×	2
Choice in μ_R and μ_F	1–4	✓	1
PDF choice	1–2	✓	1
Color reconnection	1.5	✓	< 1
Parton shower	1–8	✓	1
WZ cross section	10–20	✓	3
WZ + heavy flavor	8	✓	1
ZZ cross section	10	✓	1
$\mathrm{t(\bar{t})X}$ bg.	10–15	✓	3
$\mathrm{X}\gamma$ background	20	✓	1
Nonprompt background	30	✓	< 1
Rare SM background	50	✓	2
Stat. unc. in nonprompt bg.	5–50	×	< 1
Stat. unc. in rare SM bg.	5–100	×	< 1
Total uncertainty			7

are listed in the fourth column of the Table 2. The largest contribution comes from the imperfect knowledge of the lepton selection efficiencies in the signal acceptance. The uncertainties in parton shower modeling and $\mathrm{t(\bar{t})X}$ background yields also form a large fraction of the total uncertainty. With respect to the earlier measurements [4, 5], the total uncertainty on the inclusive cross section is reduced by about 40%. While this improvement is primarily due to a better lepton selection procedure and the detailed studies of its performance in simulation, improved estimation of the trigger and b tagging selection efficiencies in simulation, and the reduced statistical uncertainty through the addition of more data also play important role. The reported result is the first experimental measurement that is more precise than the state-of-the-art theoretical calculations for $\mathrm{t\bar{t}Z}$ production at NLO in QCD.

The signal-enriched region is defined by selecting events with $N_b \geq 1$ and $N_j \geq 3$ (2) for the three (four)-lepton channels. In this region, the signal purity is about 65%. In Fig. 4 several kinematic distributions in this signal-enriched region are shown. The sum of the signal and background predictions are found to describe the data well. The event yields are shown in Table 4.

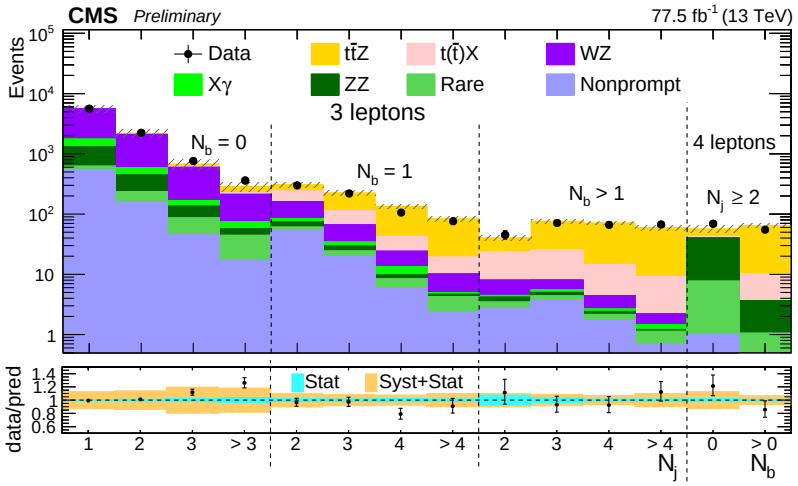


Figure 3: Observed data as a function of N_j and N_b for events with 3 and 4 leptons, compared to the simulated signal and background yields, as obtained from the fit. The hatched band shows the total uncertainty associated with the signal and background predictions, as obtained from the fit.

Table 3: Measured $t\bar{t}Z$ signal strengths for events with 3 and 4 leptons, and the value from the combined fit.

Fit configuration	Measured signal strength
3ℓ	$1.24\ ^{+0.07}_{-0.07}(\text{stat})\ ^{+0.08}_{-0.08}(\text{syst})$
4ℓ	$1.15\ ^{+0.18}_{-0.17}(\text{stat})\ ^{+0.10}_{-0.08}(\text{syst})$
Combined	$1.19\ ^{+0.07}_{-0.06}(\text{stat})\ ^{+0.08}_{-0.07}(\text{syst})$

Table 4: Predicted and observed yields, and total uncertainties in the signal-enriched sample of events.

Process	$(\mu)\mu\mu\mu$	$(\mu)\mu\mu$ e	$(\mu/\text{e})\mu$ ee	(e)eee	Total
$t\bar{t}Z$	152 ± 8	129 ± 7	118 ± 6	82 ± 4	481 ± 24
$t\bar{t}H$	4.0 ± 0.5	3.5 ± 0.4	3.2 ± 0.4	2.1 ± 0.3	12.7 ± 1.5
$t(\bar{t})X$	33.3 ± 4.1	27.4 ± 3.4	$23. \pm 2.9$	17.9 ± 2.2	102 ± 12
WZ	17.1 ± 4.6	14.7 ± 4.1	10.0 ± 2.8	10.9 ± 3.0	52.8 ± 14.2
$X\gamma$	1.6 ± 1.6	2.1 ± 2.5	0.6 ± 0.6	4.5 ± 1.6	8.8 ± 3.7
ZZ	2.8 ± 0.4	2.7 ± 0.4	2.6 ± 0.3	2.2 ± 0.3	10.3 ± 1.3
Rare	3.9 ± 2.0	2.9 ± 1.5	2.6 ± 1.3	2.0 ± 1.0	11.3 ± 5.7
Nonprompt	7.3 ± 3.0	11.2 ± 4.2	7.2 ± 3.0	8.9 ± 3.6	34.5 ± 13.1
Total	222 ± 13	194 ± 12	168 ± 9	130 ± 8	713 ± 41
Observed	192	175	152	141	660

6.2 Differential cross section measurement

The differential cross section is measured as a function of the transverse momentum $p_T(Z)$ of the Z boson and $\cos\theta^*_Z$, the cosine of the angle between the Z boson and the negatively charged lepton from the Z boson decay, as measured in the Z rest frame. The momentum of the Z boson is defined after QCD radiation has been simulated. The differential cross section is defined in the same phase space as the inclusive cross section reported above, i.e., in the phase space

10

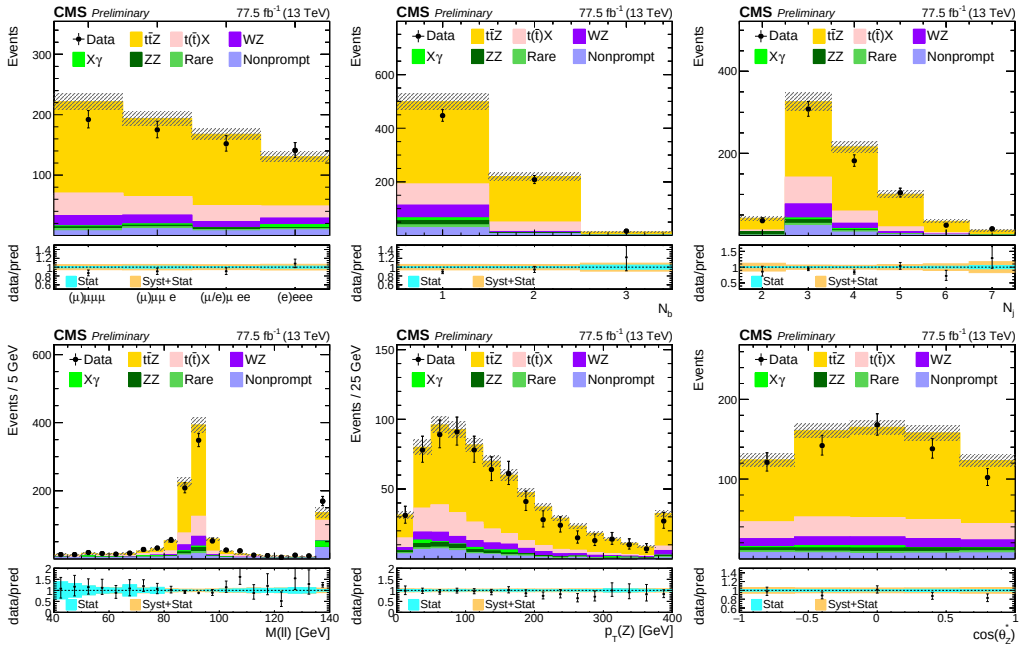


Figure 4: Observed data in a $\mathrm{t\bar{t}Z}$ dominated region, compared to signal and background yields, as obtained from the fit. Event distributions are shown as a function of the lepton flavor (upper left), N_b (upper middle), N_j (upper right), dilepton invariant mass $m(\ell\ell)$ (lower left), $p_\mathrm{T}(\mathrm{Z})$ (lower middle), and $\cos\theta^*_\mathrm{Z}$ (lower right). The hatched band shows the total uncertainty associated with the signal and background predictions, as obtained from the fit.

where the top quark pair is produced in association with two leptons with an invariant mass of more than 10 GeV, correcting for detector efficiencies and acceptances, as well as for the branching fraction for a Z boson to decay into a pair of muons or electrons.

The measurement of the differential cross section is performed in a signal-enriched sample of events defined by requiring exactly three identified leptons, $N_\mathrm{b} \geq 1$, and $N_\mathrm{j} \geq 3$. The corresponding post-fit distributions are shown in Fig. 4. Since the data sets under study are statistically limited, a rather coarse binning is chosen, with four bins in each distribution. The number of signal events in each reconstructed bin is determined by subtracting the expected number of background events from the number of events in the data.

The simulated $\mathrm{t\bar{t}Z}$ Monte Carlo are used to construct a response matrix that takes into account both detector response and acceptance corrections. The same corrections, scale factors, and uncertainties as used in the inclusive cross section are applied. Since the resolution of the lepton momenta is very good, the fraction of events migrating from one bin to another is extremely small. In all bins, the purity, defined as the fraction of reconstructed events that originate from the same bin and the stability, defined as the fraction of generated events that are reconstructed in the same bin, are larger than 94%. Under such conditions, matrix inversion without regularization is expected to provide unbiased and stable results. In this analysis, the `TUnfold` package [66], is used to obtain the results for the two measured observables.

For each theoretical uncertainty in the signal sample, the response matrix is modified accordingly and the unfolding procedure is repeated. Uncertainties due to the background expectation are accounted for by the corresponding variation of the number of subtracted background events. Experimental uncertainties due to detector response and efficiency, such as lepton iden-

tification, jet energy scale, and b tagging uncertainties, are applied as a function of the reconstructed observable. For the latter uncertainties, the unfolding is performed using the same response matrix as for the nominal result, and varying the input data within their uncertainties. This choice is made in order to minimize possible contributions from numerical effects in the matrix inversion.

In Fig. 5, the measured absolute and normalized cross sections, as obtained from the unfolding procedure described above, are shown as function of $p_{\mathrm{T}}(\mathrm{Z})$ and $\cos\theta^*_{\mathrm{Z}}$. Also shown is the prediction from the Monte Carlo generator MADGRAPH5_aMC@NLO and its uncertainty due to scale variations. The scale variations affect the normalization of the prediction but not its shape. In agreement with the inclusive cross section measurement reported above, the absolute differential cross section is measured to be slightly higher than the prediction, while the shape is described well.

6.3 Search for anomalous couplings and effective field theory interpretation

The role of the top quark in many BSM models [6–12] makes its interactions, in particular, the electroweak gauge couplings, sensitive probes that can be exploited by interpreting the differential $\mathrm{t\bar{t}Z}$ cross section in models with modified t-Z interactions. Going beyond earlier work, where the inclusive cross section measurement [4] was used, we first consider an anomalous-coupling Lagrangian [67]

$$\mathcal{L} = e\bar{u}(p_t)\left[\gamma^\mu\left(C_{1,V} + \gamma_5 C_{1,A}\right) + \frac{\mathrm{i}\sigma^{\mu\nu}q_\nu}{M_{\mathrm{Z}}}\left(C_{2,V} + \mathrm{i}\gamma_5 C_{2,A}\right)\right]v(p_{\bar{t}})\,\mathrm{Z}_\mu, \tag{1}$$

which contains the neutral vector and axial-vector current couplings, $C_{1,V}$ and $C_{1,A}$, as well as the weak magnetic and electric dipole interaction couplings, $C_{2,V}$ and $C_{2,A}$. In total, there are four real parameters. The current couplings are exactly predicted by the SM as $C^{\mathrm{SM}}_{1,V} = \frac{I^f_{3,q} - 2Q_f \sin^2\theta_W}{2\sin\theta_W\cos\theta_W} = 0.2448(52)$ and $C^{\mathrm{SM}}_{1,A} = \frac{-I^f_{3,q}}{2\sin\theta_W\cos\theta_W} = -0.6012(14)$, where θ_W is the Weinberg angle, and Q_f and $I^f_{3,q}$ label the charge and the third component of the isospin of the SM fermions [68]. The dipole moments, moreover, are generated only radiatively in the SM, and their small numerical values, which are well below 10^{-3} [6, 69, 70], allow stringent tests. Beyond the transverse momentum of the Z boson, several observables have been considered that are sensitive to anomalous electroweak interactions of the top quark [71]. Among them, $\cos\theta^*_{\mathrm{Z}}$ has a high experimental resolution and provides the best discriminating power when compared to a comprehensive set of alternative choices calculated using the reconstructed leptons, jets and b-tagged jets.

An alternative interpretation is given in the context of SM EFT in the Warsaw basis [14] formed by 59 independent Wilson coefficients of mass dimension 6. Among them, 15 are relevant for top quark interactions [72], which in general have a large impact on processes other than $\mathrm{t\bar{t}Z}$. Anomalous interactions between the top quark and the gluon (chromomagnetic and chromoelectric dipole moment interactions) are tightly constrained by the $\mathrm{t\bar{t}}$+jets final state [73]. Similarly, the modification of the Wtb vertex is best constrained by measurements of the W helicity fractions in top quark pair production [74] and in the t-channel single top quark production [75]. It is thus appropriate to separately consider the operators that induce anomalous interactions of the top quark with the remaining neutral gauge bosons, the Z boson and the photon. In the parametrization adopted here [15], their Wilson coefficients are c_{tZ}, $c^{[I]}_{tZ}$, $c_{\phi t}$, and $c^{-}_{\phi Q}$. The former two induce electroweak dipole moments, while the latter two induce anomalous neutral-current interactions. These Wilson coefficients are the main focus of this work,

12

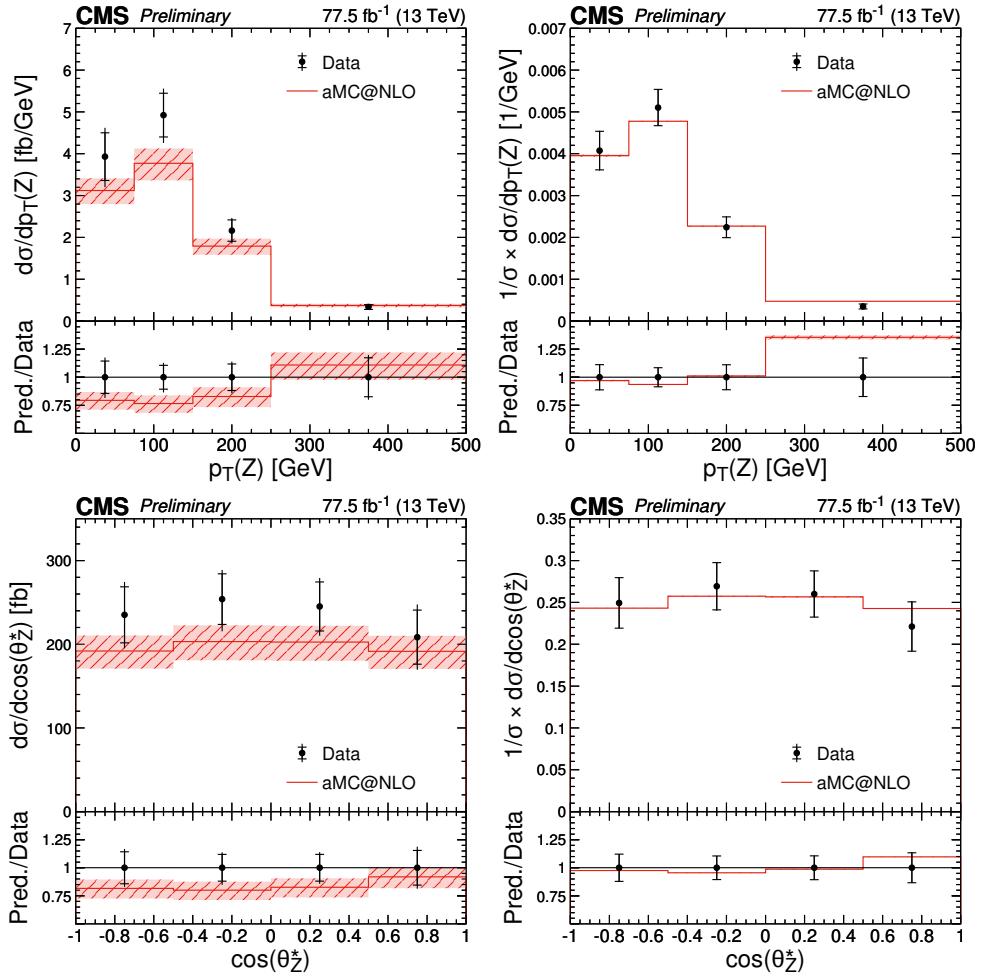


Figure 5: Measured differential $t\bar{t}Z$ production cross sections in the full phase space as a function of the transverse momentum $p_T(Z)$ of the Z boson (top row), and $\cos\theta^*_Z$, as defined in the text (bottom row). Shown are the absolute (left) and normalized (right) cross sections. The data are represented by the points. The inner (outer) vertical bars indicate the statistical (total) uncertainties, respectively. The histogram shows the prediction from the MADGRAPH5_aMC@NLO Monte Carlo simulation, and the hatched band indicates the theory uncertainties in the prediction. The lower planels display the ratio between prediction and measurement.

which are combined as

$$c_{tZ} = \mathrm{Re}\left(-\sin\theta_W C_{uB}^{(33)} + \cos\theta_W C_{uW}^{(33)}\right) \quad (2)$$

$$c_{tZ}^{[I]} = \mathrm{Im}\left(-\sin\theta_W C_{uB}^{(33)} + \cos\theta_W C_{uW}^{(33)}\right) \quad (3)$$

$$c_{\phi t} = C_{\phi t} = C_{\phi u}^{(33)} \quad (4)$$

$$c_{\phi Q}^{-} = C_{\phi Q} = C_{\phi q}^{1(33)} - C_{\phi q}^{3(33)}, \quad (5)$$

where θ_W is the weak mixing angle and the Wilson coefficients in the Warsaw basis are denoted by $C_{uB}^{(33)}$, $C_{uW}^{(33)}$, $C_{\phi u}^{(33)}$, $C_{\phi q}^{1(33)}$, and $C_{\phi q}^{3(33)}$ as defined in Ref. [15]. The constraints $C_{\phi q}^{3(33)} = 0$ and

$C_{uW}^{(33)} = 0$ ensure a Wtb vertex according to the SM.

Based on the expected sensitivity, we choose the following signal regions in the three- and four-lepton channel. In the three-lepton channel, there are 12 signal regions defined by the three $p_T(Z)$ thresholds 0, 200, and 400 GeV and four thresholds on $\cos\theta_Z^*$ at -1, -0.6, 0.6, and 1. In the four-lepton channel, the predicted event yields are lower, leading to an optimal choice of only three bins defined in terms of $p_T(Z)$ with thresholds at 0, 100, and 200 GeV. The jet multiplicity requirement is relaxed to $N_j \geq 1$. Next, 12 control regions in the three-lepton channel are defined by requiring $N_b = 0$ and $N_j \geq 1$, but otherwise reproducing the three-lepton signal selections. The three-lepton control regions guarantee a pure selection of the main WZ background. In order to also constrain the leading ZZ background of the four-lepton channel, we add three more control regions with $N_j \geq 1$ and $N_b \geq 0$ and require that there are two pairs of opposite-sign same-flavor leptons consistent with the Z boson mass in a window of 15 GeV. A summary of the signal and control regions is given in Table 5.

Predictions for signal yields with nonzero values of anomalous couplings or Wilson coefficients are obtained by simulating large LO samples in the respective model on a fine grid in parameter space, including the SM configuration. Then, the two-dimensional generator-level distributions of $p_T(Z)$ and $\cos\theta_Z^*$ for the BSM and the SM parameter points are used to define the reweighting of the nominal NLO $t\bar{t}Z$ sample. The result of the reweighting procedure is tested on a coarse grid in BSM parameter space, where BSM samples are produced and reconstructed. The differences between the full event reconstruction and the reweighting procedure are found to be negligible for all distributions considered in this work. The theoretical uncertainties in the predicted BSM yields are scaled accordingly.

From the predicted yields and the uncertainties, we construct a binned likelihood function $L(\theta)$ as a product of Poisson probabilities, where θ labels the set of nuisance parameters. The test statistic is the profiled likelihood ratio $q = -2\log(L(\hat{\theta})/L(\hat{\theta}_{max}))$ where $\hat{\theta}$ is the set of nuisance parameters maximizing the likelihood function at the BSM point, while $\hat{\theta}_{max}$ is the maximum of all $\hat{\theta}$ in the BSM parameter plane.

In Fig. 6 we show the best-fit result in the plane spanned by $c_{\phi t}$ and $c_{\phi Q}^-$ using the regions in Table 5. In Fig. 7 we show the log-likelihood scan in the two-dimensional (2D) planes spanned by $c_{\phi t}$ and $c_{\phi Q}^-$, as well as c_{tZ} and $c_{tZ}^{[I]}$. Consistent with the measurement of the cross section, the SM value is close to the contour of 95% confidence level (CL). The complementary scan in the plane spanned by the anomalous current interactions $C_{1,V}$ and $C_{1,A}$, as well as the anomalous dipole interactions $C_{2,V}$ and $C_{2,A}$, is shown in Fig. 8.

Finally, in Figs. 9 and 10 we show the one-dimensional (1D) scan where all other coupling parameters are set to their SM values. The corresponding one-dimensional exclusion intervals at 68% and 95% CL are listed in Table 6. A comparison of the observed 95% confidence intervals with earlier measurements is shown in Fig. 11.

14

Table 5: Definition of the signal and control regions.

N_ℓ	N_j	N_b	N_Z	$p_T(Z)$ (GeV)	$-1 \leq \cos(\theta^*) < -0.6$	$-0.6 \leq \cos(\theta^*) < 0.6$	$0.6 \leq \cos(\theta^*)$
3	≥ 3	≥ 1	1	0–100	SR1	SR2	SR3
				100–200	SR4	SR5	SR6
				200–400	SR7	SR8	SR9
				≥ 400	SR10	SR11	SR12
4	≥ 1	≥ 1	1	0–100	SR13		
				100–200	SR14		
				≥ 200	SR15		
3	≥ 1	0	1	0–100	CR1	CR2	CR3
				100–200	CR4	CR5	CR6
				200–400	CR7	CR8	CR9
				≥ 400	CR10	CR11	CR12
4	≥ 1	≥ 0	2	0–100	CR13		
				100–200	CR14		
				≥ 200	CR15		

Table 6: Expected and observed 68% and 95% CL intervals from this measurement for the listed Wilson coefficients. Constraints from a previous CMS measurement [4] and indirect constraints from precision electroweak data [77] are shown for comparison.

Coefficient	Expected		Observed		Previous CMS constraints		Indirect constraints
	68% CL	95% CL	68% CL	95% CL	Exp, 95% CL	Obs, 95% CL	68% CL
c_{tZ}/Λ^2	$[-0.7, 0.7]$	$[-1.1, 1.1]$	$[-0.8, 0.5]$	$[-1.1, 1.1]$	$[-2.0, 2.0]$	$[-2.6, 2.6]$	$[-4.7, 0.2]$
$c_{tZ}^{[I]}/\Lambda^2$	$[-0.7, 0.7]$	$[-1.1, 1.1]$	$[-0.8, 1.0]$	$[-1.2, 1.2]$	–	–	–
$c_{\phi t}/\Lambda^2$	$[-1.6, 1.4]$	$[-3.4, 2.8]$	$[2.2, 4.7]$	$[0.7, 5.9]$	$[-20.2, 4.0]$	$[-22.2, -13.0]$ $[-3.2, 6.0]$	$[-0.1, 3.7]$
$c_{\phi Q}^{-}/\Lambda^2$	$[-1.1, 1.1]$	$[-2.1, 2.2]$	$[-3.0, -1.0]$	$[-4.0, 0.0]$	–	–	$[-4.7, 0.7]$

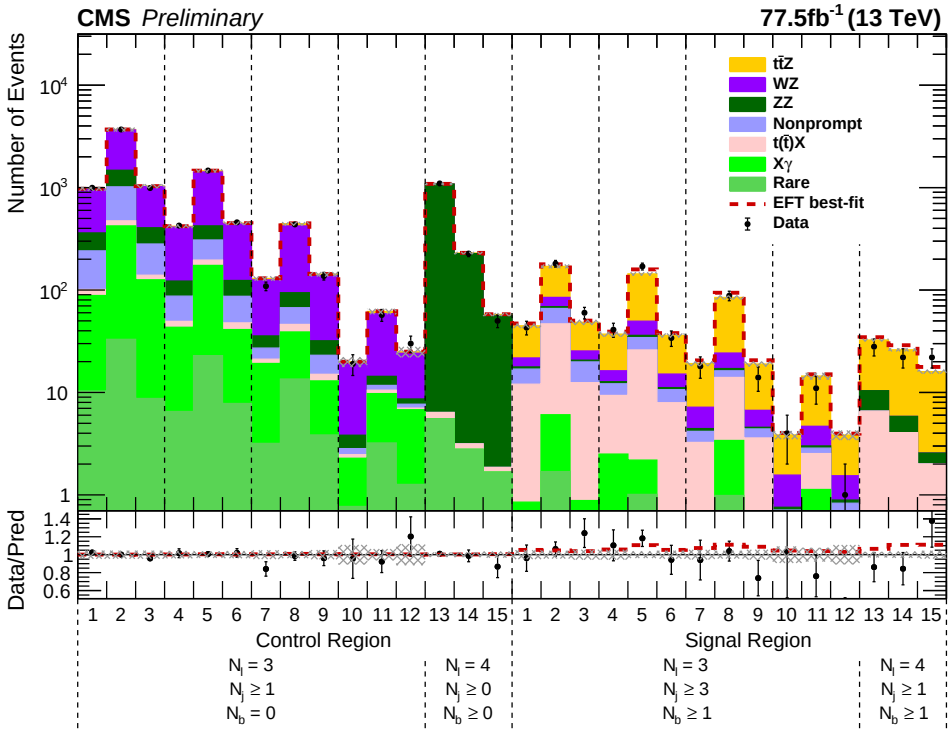


Figure 6: Predicted and observed yields (post-fit) for the combined 2016 and 2017 data sets in the control and signal regions. In the $N_\ell = 3$ control and signal regions (bins 1–12), each of the 4 $p_T(Z)$ categories is further split into 3 $\cos\theta^*_Z$ bins. Due to the lower expected event count in the $N_\ell = 4$ signal regions (bins 13–15) no categorization in terms of $\cos\theta^*_Z$ is applied. The red dashed line shows the best-fit point to the observed result in one of the EFT planes.

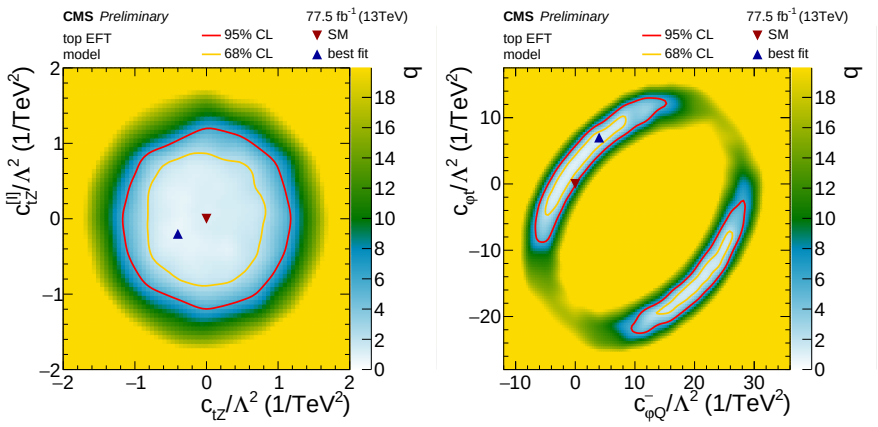


Figure 7: Results of scans in two 2D planes in the EFT interpretation. The color map reflects the negative log-likelihood ratio q w.r.t the best-fit value. The yellow and red dashed lines indicate one and two standard deviations from the best-fit value.

16

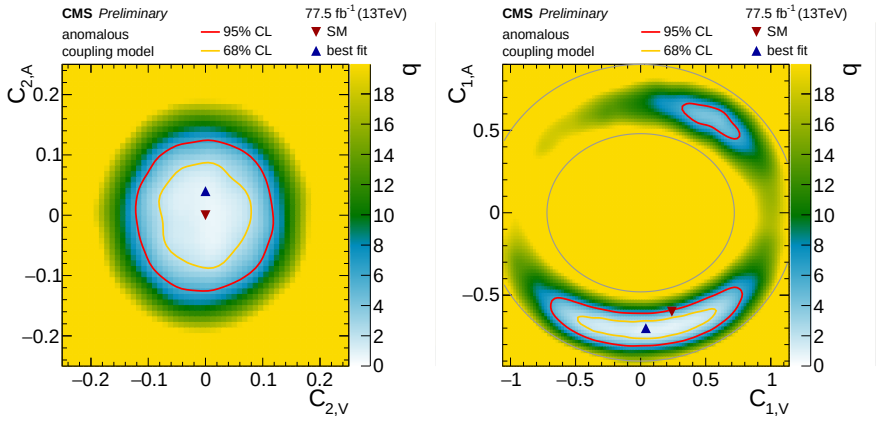


Figure 8: Results of scans in the electroweak dipole moment plane (left) and axial-vector and vector current coupling plane (right). The color map reflects the negative log-likelihood ratio w.r.t the best-fit value. The yellow and red dashed lines indicate one and two standard deviations from the best-fit value. The area between the gray ellipses in the axial-vector and vector current coupling plane corresponds to the observed 68% C.L. area from a previous CMS result [76].

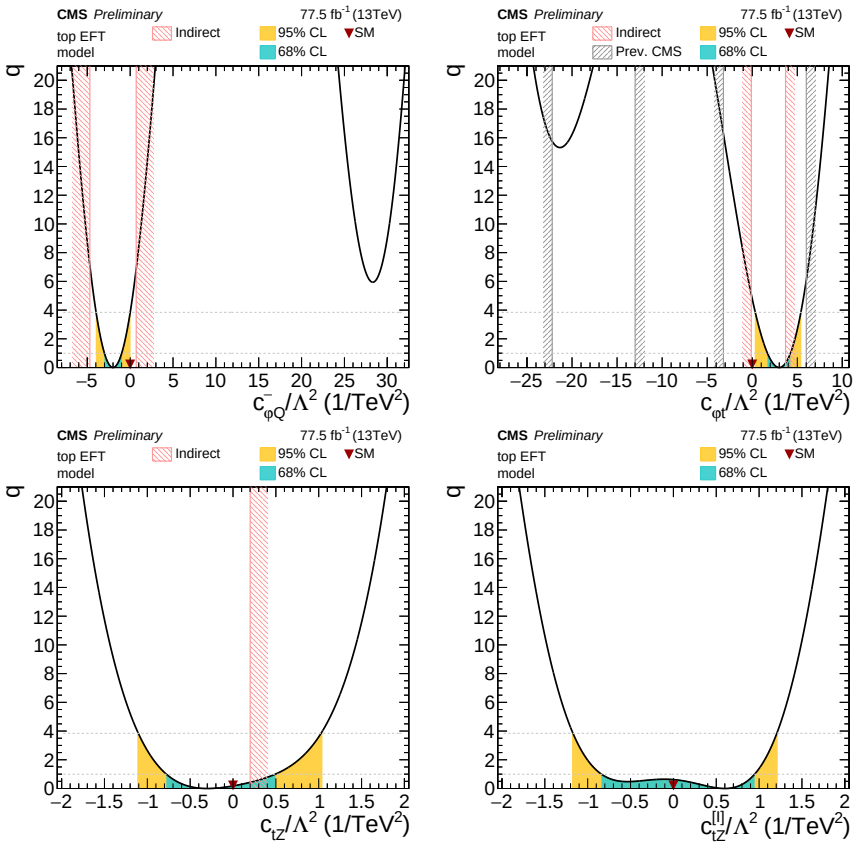


Figure 9: Log-likelihood ratios for 1D scans of Wilson coefficients. Wilson coefficients that are not shown on the respective plot are set to 0. The cyan and orange colored areas correspond to the 68% and 95% confidence level intervals around the best-fit value, respectively. The red triangle indicates the SM value. Previously excluded regions at 95% CL [4] (if available) are indicated by the gray hatched band. Indirect constraints from [77] are shown as light red hatched band.

18

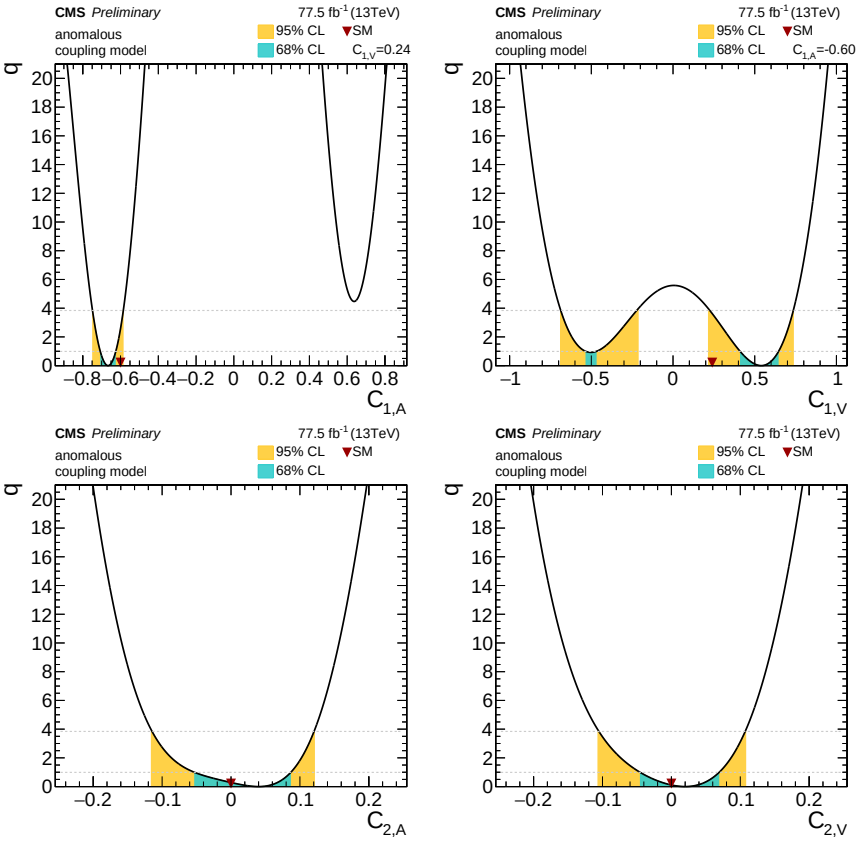


Figure 10: Log-likelihood ratios for 1D scans of anomalous couplings. $C_{1,V} = 0.24$ (SM value) for the scan of $C_{1,A}$ (top left) and $C_{1,A} = -0.60$ (SM value) for the scan of $C_{1,V}$ (top right). The cyan and orange colored areas correspond to the 68% and 95% confidence level intervals around the best-fit value, respectively. The red triangles indicate the SM values.

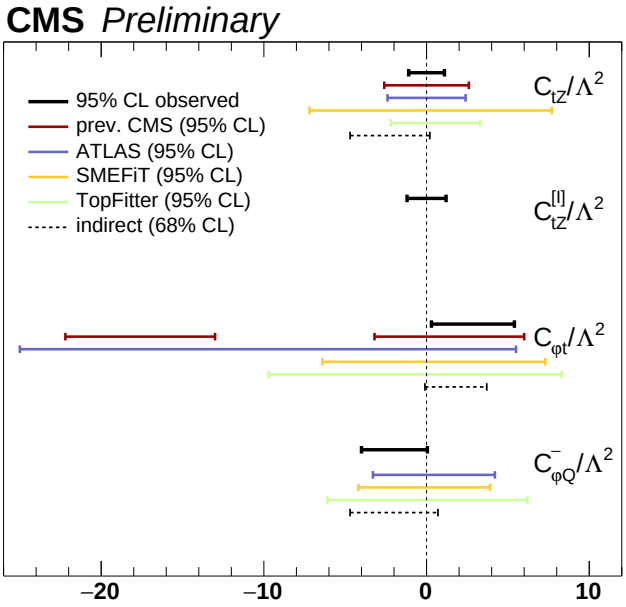


Figure 11: Comparison of the observed 95% confidence level intervals (solid black) with the previous CMS result based on the inclusive cross section measurement [4] (red), the most recent ATLAS result [5] (blue), direct limits from the SMEFiT framework [78] (orange) and the TopFitter collaboration [79] (green), as well as indirect limits from electroweak data [77] (dashed black) on the Wilson coefficients.

7 Summary

A measurement has been presented of top quark pair production in association with a Z boson using a data sample of proton-proton collisions at $\sqrt{s}$ = 13 TeV, corresponding to an integrated luminosity of 77.5 fb^{-1}, collected by the CMS detector. The analysis is performed in the three- and four-lepton final states using analysis categories defined with jet and b jet multiplicities. Data samples enriched in background processes are used to validate predictions, as well as to constrain their uncertainties. Thanks to the larger data set used and reduced systematic uncertainties such as those associated with the lepton identification, the precision on the measured cross section is substantially improved with respect to previous measurements reported in Refs. [4, 5]. The measured inclusive cross section is $\sigma(\text{t}\bar{\text{t}}\text{Z}) = 1.00\,^{+0.06}_{-0.05}\,(\text{stat})\,^{+0.07}_{-0.06}\,(\text{syst})$ pb, in agreement with the standard model prediction. This is not only the most precise measurement today, but also the first measurement with better precision compared to the theoretical prediction at NLO in QCD. Furthermore, absolute and normalized differential cross sections for the transverse momentum of the Z boson as well as $\cos\theta^*_{\text{Z}}$ are presented for the first time. The SM predictions at NLO are found to be in good agreement with the measured inclusive and differential cross sections. The measurement is also interpreted in terms of anomalous interactions of the t quark with the Z boson. Confidence intervals for anomalous vector and axial-vector current couplings and dipole moment interactions are presented. Wilson coefficients in the top quark effective field theory are similarly constrained.

References

[1] D. de Florian et al., "Handbook of LHC Higgs cross sections: 4. deciphering the nature of the Higgs sector", CERN Report CERN-2017-002-M, 2016. doi:10.23731/CYRM-2017-002, arXiv:1610.07922.

[2] O. Bessidskaia Bylund et al., "Probing top quark neutral couplings in the Standard Model Effective Field Theory at NLO in QCD", *JHEP* **05** (2016) 052, doi:10.1007/JHEP05(2016)052, arXiv:1601.08193.

[3] C. Englert, R. Kogler, H. Schulz, and M. Spannowsky, "Higgs coupling measurements at the LHC", *Eur. Phys. J. C* **76** (2016) 393, doi:10.1140/epjc/s10052-016-4227-1, arXiv:1511.05170.

[4] CMS Collaboration, "Measurement of the cross section for top quark pair production in association with a W or Z boson in proton-proton collisions at $\sqrt{s} = 13$ TeV", *JHEP* **08** (2018) 011, doi:10.1007/JHEP08(2018)011, arXiv:1711.02547.

[5] ATLAS Collaboration, "Measurement of the $t\bar{t}Z$ and $t\bar{t}W$ cross sections in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector", arXiv:1901.03584.

[6] W. Hollik et al., "Top dipole form-factors and loop induced CP violation in supersymmetry", *Nucl. Phys.* **B551** (1999) 3, doi:10.1016/S0550-3213(99)00396-X, 10.1016/S0550-3213(99)00201-1, arXiv:hep-ph/9812298. [Erratum: Nucl. Phys.B557,407(1999)].

[7] K. Agashe, G. Perez, and A. Soni, "Collider signals of top quark flavor violation from a warped extra dimension", *Phys. Rev.* **D75** (2007) 015002, doi:10.1103/PhysRevD.75.015002.

[8] A. L. Kagan, G. Perez, T. Volansky, and J. Zupan, "General minimal flavor violation", *Phys. Rev.* **D80** (2009) 076002, doi:10.1103/PhysRevD.80.076002.

[9] T. Ibrahim and P. Nath, "The top quark electric dipole moment in an MSSM extension with vector like multiplets", *Phys. Rev.* **D82** (2010) 055001, doi:10.1103/PhysRevD.82.055001.

[10] T. Ibrahim and P. Nath, "The Chromoelectric dipole moment of the top quark in models with vector like multiplets", *Phys. Rev.* **D84** (2011) 015003, doi:10.1103/PhysRevD.84.015003.

[11] C. Grojean, O. Matsedonskyi, and G. Panico, "Light top partners and precision physics", *JHEP* **10** (2013) 160, doi:10.1007/JHEP10(2013)160.

[12] F. Richard, "Can LHC observe an anomaly in ttZ production?", arXiv:1304.3594.

[13] J. A. Aguilar-Saavedra, "A Minimal set of top anomalous couplings", *Nucl. Phys.* **B812** (2009) 181–204, doi:10.1016/j.nuclphysb.2008.12.012, arXiv:0811.3842.

[14] B. Grzadkowski, M. Iskrzynski, M. Misiak, and J. Rosiek, "Dimension-Six Terms in the Standard Model Lagrangian", *JHEP* **10** (2010) 085, doi:10.1007/JHEP10(2010)085, arXiv:1008.4884.

[15] J. A. Aguilar Saavedra et al., "Interpreting top-quark LHC measurements in the standard-model effective field theory", arXiv:1802.07237.

[16] J. Alwall et al., "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations", *JHEP* **07** (2014) 079, doi:10.1007/JHEP07(2014)079, arXiv:1405.0301.

[17] S. Alioli, P. Nason, C. Oleari, and E. Re, "A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX", *JHEP* **06** (2010) 043, doi:10.1007/JHEP06(2010)043, arXiv:1002.2581.

[18] NNPDF Collaboration, "Parton distributions for the LHC Run II", *JHEP* **04** (2015) 040, doi:10.1007/JHEP04(2015)040, arXiv:1410.8849.

[19] NNPDF Collaboration, "Parton distributions from high-precision collider data", *Eur. Phys. J.* **C77** (2017), no. 10, 663, doi:10.1140/epjc/s10052-017-5199-5, arXiv:1706.00428.

[20] T. Sjöstrand, S. Mrenna, and P. Z. Skands, "A brief introduction to PYTHIA 8.1", *Comput. Phys. Commun.* **178** (2008) 852, doi:10.1016/j.cpc.2008.01.036, arXiv:0710.3820.

[21] T. Sjöstrand et al., "An introduction to PYTHIA 8.2", *Comput. Phys. Commun.* **191** (2015) 159, doi:10.1016/j.cpc.2015.01.024, arXiv:1410.3012.

[22] P. Skands, S. Carrazza, and J. Rojo, "Tuning PYTHIA 8.1: the Monash 2013 tune", *Eur. Phys. J. C* **74** (2014) 3024, doi:10.1140/epjc/s10052-014-3024-y, arXiv:1404.5630.

[23] CMS Collaboration, "Event generator tunes obtained from underlying event and multiparton scattering measurements", *Eur. Phys. J. C* **76** (2016) 155, doi:10.1140/epjc/s10052-016-3988-x, arXiv:1512.00815.

[24] CMS Collaboration, "Investigations of the impact of the parton shower tuning in Pythia 8 in the modelling of $t\bar{t}$ at $\sqrt{s} = 8$ and 13 TeV", Technical Report CMS-PAS-TOP-16-021, CERN, Geneva, 2016.

[25] J. Alwall et al., "Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions", *Eur. Phys. J. C* **53** (2008) 473, doi:10.1140/epjc/s10052-007-0490-5, arXiv:0706.2569.

[26] R. Frederix and S. Frixione, "Merging meets matching in MC@NLO", *JHEP* **12** (2012) 061, doi:10.1007/JHEP12(2012)061, arXiv:1209.6215.

[27] J. Butterworth et al., "PDF4LHC recommendations for LHC Run II", *J. Phys.* **G43** (2016) 023001, doi:10.1088/0954-3899/43/2/023001, arXiv:1510.03865.

[28] F. Caola, K. Melnikov, R. Röntsch, and L. Tancredi, "QCD corrections to ZZ production in gluon fusion at the LHC", *Phys. Rev. D* **92** (2015) 094028, doi:10.1103/PhysRevD.92.094028, arXiv:1509.06734.

[29] J. M. Campbell and R. K. Ellis, "MCFM for the Tevatron and the LHC", *Nucl. Phys. Proc. Suppl.* **205-206** (2010) 10, doi:10.1016/j.nuclphysbps.2010.08.011, arXiv:1007.3492.

[30] S. Bolognesi et al., "On the spin and parity of a single-produced resonance at the LHC", *Phys. Rev.* **D86** (2012) 095031, doi:10.1103/PhysRevD.86.095031, arXiv:1208.4018.

[31] F. Cascioli et al., "ZZ production at hadron colliders in NNLO QCD", *Phys. Lett. B* **735** (2014) 311, doi:10.1016/j.physletb.2014.06.056, arXiv:1405.2219.

[32] T. Melia, P. Nason, R. Rontsch, and G. Zanderighi, "W^+W^-, WZ and ZZ production in the POWHEG BOX", *JHEP* **11** (2011) 078, doi:10.1007/JHEP11(2011)078, arXiv:1107.5051.

[33] P. Nason and G. Zanderighi, "W^+W^- , WZ and ZZ production in the POWHEG-BOX-V2", *Eur. Phys. J. C* **74** (2014) 2702, doi:10.1140/epjc/s10052-013-2702-5, arXiv:1311.1365.

[34] G. Luisoni, P. Nason, C. Oleari, and F. Tramontano, "$HW^{\pm}$/HZ + 0 and 1 jet at NLO with the POWHEG BOX interfaced to GoSam and their merging within MiNLO", *JHEP* **10** (2013) 083, doi:10.1007/JHEP10(2013)083, arXiv:1306.2542.

[35] H. B. Hartanto, B. Jager, L. Reina, and D. Wackeroth, "Higgs boson production in association with top quarks in the POWHEG BOX", *Phys. Rev. D* **91** (2015) 094003, doi:10.1103/PhysRevD.91.094003, arXiv:1501.04498.

[36] M. Czakon and A. Mitov, "Top++: A program for the calculation of the top-pair cross section at hadron colliders", *Comput. Phys. Commun.* **185** (2014) 2930, doi:10.1016/j.cpc.2014.06.021, arXiv:1112.5675.

[37] GEANT4 Collaboration, "GEANT4 — a simulation toolkit", *Nucl. Instrum. Meth. A* **506** (2003) 250, doi:10.1016/S0168-9002(03)01368-8.

[38] CMS Collaboration, "Particle-flow reconstruction and global event description with the CMS detector", *JINST* **12** (2017) P10003, doi:10.1088/1748-0221/12/10/P10003, arXiv:1706.04965.

[39] M. Cacciari, G. P. Salam, and G. Soyez, "The anti-k_t jet clustering algorithm", *JHEP* **04** (2008) 063, doi:10.1088/1126-6708/2008/04/063, arXiv:0802.1189.

[40] CMS Collaboration, "Pileup removal algorithms", CMS Physics Analysis Summary CMS-PAS-JME-14-001, 2014.

[41] CMS Collaboration, "Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV", *JINST* **13** (2018), no. 05, P05011, doi:10.1088/1748-0221/13/05/P05011, arXiv:1712.07158.

[42] CMS Collaboration, "Observation of $\mathrm{t\bar{t}H}$ production", *Phys. Rev. Lett.* **120** (2018), no. 23, 231801, doi:10.1103/PhysRevLett.120.231801, arXiv:1804.02610.

[43] A. Hocker et al., "TMVA - Toolkit for multivariate data analysis", arXiv:physics/0703039.

[44] J. Campbell, R. K. Ellis, and R. Röntsch, "Single top production in association with a Z boson at the LHC", *Phys. Rev. D* **87** (2013) 114006, doi:10.1103/PhysRevD.87.114006, arXiv:1302.3856.

[45] S. Frixione et al., "Electroweak and QCD corrections to top-pair hadroproduction in association with heavy bosons", *JHEP* **06** (2015) 184, doi:10.1007/JHEP06(2015)184, arXiv:1504.03446.

[46] CMS Collaboration, "Observation of single top quark production in association with a Z boson in proton-proton collisions at $\sqrt{s} = 13$ TeV", *Submitted to: Phys. Rev. Lett.* (2018) arXiv:1812.05900.

[47] CMS Collaboration, "Measurement of differential cross sections for Z boson pair production in association with jets at $\sqrt{s} = 8$ and 13 TeV", arXiv:1806.11073.

[48] D. T. Nhung, L. D. Ninh, and M. M. Weber, "NLO corrections to WWZ production at the LHC", *JHEP* **12** (2013) 096, doi:10.1007/JHEP12(2013)096, arXiv:1307.7403.

[49] CMS Collaboration, "Measurement of differential cross sections for Z boson production in association with jets in proton-proton collisions at $\sqrt{s} = 13$ TeV", *Eur. Phys. J.* **C78** (2018), no. 11, 965, doi:10.1140/epjc/s10052-018-6373-0, arXiv:1804.05252.

[50] CMS Collaboration, "CMS luminosity measurement for the 2016 data taking period", CMS Physics Analysis Summary CMS-PAS-LUM-17-001, 2017.

[51] ATLAS Collaboration, "Measurement of the inelastic proton-proton cross section at $\sqrt{s} = 13$ TeV with the ATLAS detector at the LHC", *Phys. Rev. Lett.* **117** (2016) 182002, doi:10.1103/PhysRevLett.117.182002, arXiv:1606.02625.

[52] CMS Collaboration, "Performance of CMS muon reconstruction in pp collision events at $\sqrt{s} = 7$ TeV", *JINST* **7** (2012) P10002, doi:10.1088/1748-0221/7/10/P10002, arXiv:1206.4071.

[53] CMS Collaboration, "Performance of electron reconstruction and selection with the CMS detector in proton-proton collisions at $\sqrt{s} = 8$ TeV", *JINST* **10** (2015) P06005, doi:10.1088/1748-0221/10/06/P06005, arXiv:1502.02701.

[54] CMS Collaboration, "Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV", *JINST* **12** (2017), no. 02, P02014, doi:10.1088/1748-0221/12/02/P02014, arXiv:1607.03663.

[55] CMS Collaboration, "Identification of b-quark jets with the CMS experiment", *JINST* **8** (2013) P04013, doi:10.1088/1748-0221/8/04/P04013, arXiv:1211.4462.

[56] CMS Collaboration, "Identification of b quark jets at the CMS experiment in the LHC Run 2", CMS Physics Analysis Summary CMS-PAS-BTV-15-001, 2016.

[57] R. D. Ball et al., "Parton distributions with LHC data", *Nucl. Phys.* **B867** (2013) 244–289, doi:10.1016/j.nuclphysb.2012.10.003, arXiv:1207.1303.

[58] A. D. Martin, W. J. Stirling, R. S. Thorne, and G. Watt, "Uncertainties on α_S in global PDF analyses and implications for predicted hadronic cross sections", *Eur. Phys. J.* **C64** (2009) 653–680, doi:10.1140/epjc/s10052-009-1164-2, arXiv:0905.3531.

[59] J. Gao et al., "CT10 next-to-next-to-leading order global analysis of QCD", *Phys. Rev.* **D89** (2014), no. 3, 033009, doi:10.1103/PhysRevD.89.033009, arXiv:1302.6246.

[60] S. Argyropoulos and T. Sjstrand, "Effects of color reconnection on $t\bar{t}$ final states at the LHC", *JHEP* **11** (2014) 043, doi:10.1007/JHEP11(2014)043, arXiv:1407.6653.

[61] J. R. Christiansen and P. Z. Skands, "String Formation Beyond Leading Colour", *JHEP* **08** (2015) 003, doi:10.1007/JHEP08(2015)003, arXiv:1505.01681.

[62] T. Junk, "Confidence level computation for combining searches with small statistics", *Nucl. Instrum. Meth. A* **434** (1999) 435, doi:10.1016/S0168-9002(99)00498-2, arXiv:hep-ex/9902006.

[63] A. L. Read, "Presentation of search results: the CL_s technique", in *Durham IPPP Workshop: Advanced Statistical Techniques in Particle Physics*, p. 2693. Durham, UK, March, 2002. [J. Phys. G 28 (2002) 2693]. doi:10.1088/0954-3899/28/10/313.

[64] ATLAS and CMS Collaborations, "Procedure for the LHC Higgs boson search combination in summer 2011", ATL-PHYS-PUB-2011-011, CMS NOTE-2011/005, 2011.

[65] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, "Asymptotic formulae for likelihood-based tests of new physics", *Eur. Phys. J. C* **71** (2011) 1554, doi:10.1140/epjc/s10052-011-1554-0, arXiv:1007.1727. [Erratum: doi:10.1140/epjc/s10052-013-2501-z].

[66] S. Schmitt, "TUnfold: an algorithm for correcting migration effects in high energy physics", *JINST* **7** (2012) T10003, doi:10.1088/1748-0221/7/10/T10003, arXiv:1205.6201.

[67] R. Röntsch and M. Schulze, "Constraining couplings of top quarks to the Z boson in $t\bar{t}$ + Z production at the LHC", *JHEP* **07** (2014) 091, doi:10.1007/JHEP09(2015)132,10.1007/JHEP07(2014)091, arXiv:1404.1005. [Erratum: JHEP09,132(2015)].

[68] Particle Data Group, "Review of particle physics", *Chin. Phys. C* **40** (2016) 100001, doi:10.1088/1674-1137/40/10/100001.

[69] J. Bernabeu, D. Comelli, L. Lavoura, and J. P. Silva, "Weak magnetic dipole moments in two Higgs doublet models", *Phys. Rev.* **D53** (1996) 5222, doi:10.1103/PhysRevD.53.5222.

[70] A. Czarnecki and B. Krause, "On the dipole moments of fermions at two loops", *Acta Phys. Polon.* **B28** (1997) 829.

[71] M. Schulze and Y. Soreq, "Pinning down electroweak dipole operators of the top quark", *Eur. Phys. J.* **C76** (2016) 8/466, doi:10.1140/epjc/s10052-016-4263-x, arXiv:1603.08911.

[72] C. Zhang and S. Willenbrock, "Effective-Field-Theory approach to top quark production and decay", *Phys. Rev.* **D83** (2011) 034006, doi:10.1103/PhysRevD.83.034006, arXiv:1008.3869.

[73] CMS Collaboration, "Measurements of $t\bar{t}$ differential cross sections in proton-proton collisions at $\sqrt{s} = 13$ TeV using events containing two leptons", arXiv:1811.06625.

[74] CMS Collaboration, "Measurement of the W boson helicity fractions in the decays of top quark pairs to lepton + jets final states produced in pp collisions at $\sqrt{s} = 8$TeV", *Phys. Lett.* **B762** (2016) 512, doi:10.1016/j.physletb.2016.10.007, arXiv:1605.09047.

[75] J. A. Aguilar-Saavedra and J. Bernabeu, "W polarisation beyond helicity fractions in top quark decays", *Nucl. Phys.* **B840** (2010) 349–378, doi:10.1016/j.nuclphysb.2010.07.012, arXiv:1005.5382.

[76] CMS Collaboration, "Observation of top quark pairs produced in association with a vector boson in pp collisions at $\sqrt{s} = 8$ TeV", *JHEP* **01** (2016) 096, doi:10.1007/JHEP01(2016)096, arXiv:1510.01131.

[77] C. Zhang, N. Greiner, and S. Willenbrock, "Constraints on non-standard top quark couplings", *Phys. Rev.* **D86** (2012) 014024, doi:10.1103/PhysRevD.86.014024, arXiv:1201.6670.

[78] N. P. Hartland et al., "A Monte Carlo global analysis of the Standard Model Effective Field Theory: the top quark sector", arXiv:1901.05965.

[79] A. Buckley et al., "Constraining top quark effective theory in the LHC Run II era", *JHEP* **04** (2016) 015, doi:10.1007/JHEP04(2016)015, arXiv:1512.03360.

8.1 TRIGGER AND OBJECT SELECTION

Contrary to the previous measurement of $t\bar{t}Z$ using only 2016 data, where only single lepton triggers with an isolation requirement were used, here we benefit from the usage of a single lepton, dilepton and triplepton trigger mix to reach full trigger selection efficiency. The efficiency with these triggers is measured in three-lepton events passing the identification criteria of the analysis. To measure the efficiency in data we use unbiased data sets, selected with the triggers requiring high p_{T}^{miss}, jet multiplicity or missing-H_{T}[1]. The trigger efficiency in simulation is measured in $t\bar{t}Z$ events produced with the MADGRAPH5_aMC@NLO generator. We observe good agreement between the efficiencies in data and simulation, as shown in Fig. 8.2.

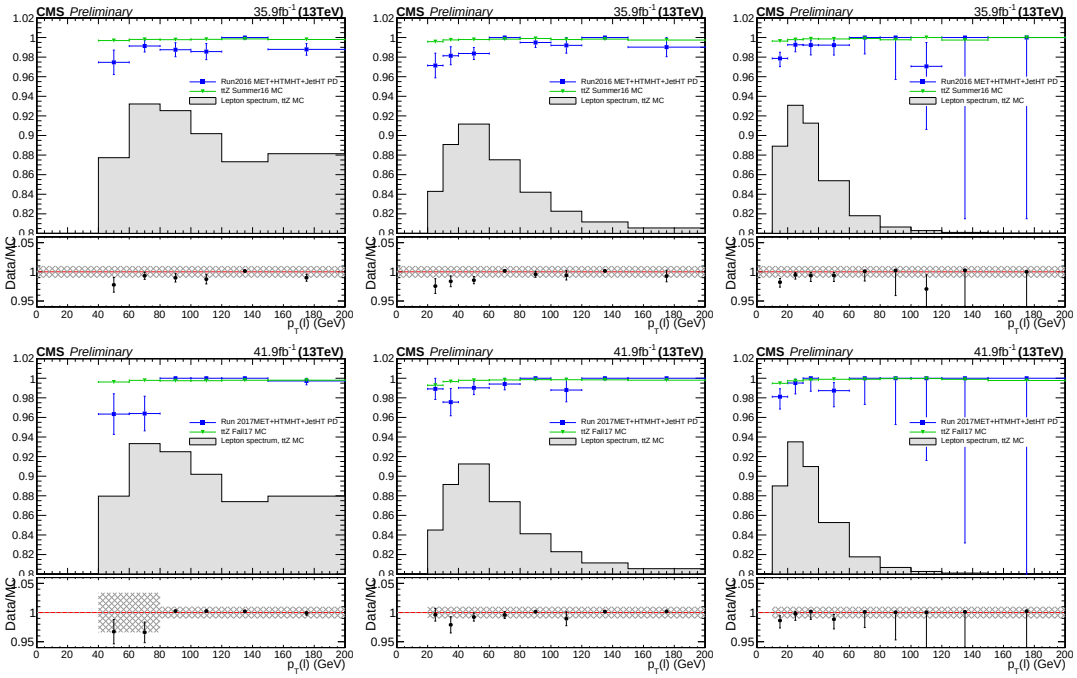


Figure 8.2: Trigger efficiencies for events with 3 selected leptons in data (blue) and simulation (green) for leading (left), sub-leading (middle) and trailing (right) lepton p_{T}. The upper plots show efficiencies measured in the 2016 data set and MC campaign, while the bottom plots depict efficiencies for 2017. The expected lepton p_{T} spectra for $t\bar{t}Z$ events are shown as filled histograms. The ratio on the bottom pad shows the data-to-simulation trigger scale factors. The figures are made by Daniel Spitzbart.

Small deviations between measured efficiencies in data and simulation are observed in events with low leading lepton p_{T}, therefore, we apply corrections to simulated events depending on this kinematical variable. For the MC simulations generated in 2016 (2017) campaign a correction factor of 0.985 (0.966) is applied to simulated events with a leading lepton p_{T} below 120 (80) GeV. We assign a 1% uncertainty uniform across all three-lepton event categories, which is increased to 3.5% for simulated events with the p_{T} of the leading lepton below 80 GeV in the 2017 data set.

Lepton identification and selection is a critical ingredient in this measurement. To achieve an effective rejection of the nonprompt leptons while keeping a high efficiency for prompt leptons, a Lepton MVA discriminator is used separately for electrons and muons. The details and performance of this selection were given in Subsection 5.2.3.

1 The variable is estimated in the same way as p_{T}^{miss} in Eq. 5.7 with the sum running over only PF jets.

Placing a stringent requirement on the Lepton MVA output, an increase of up to 15% in prompt lepton selection efficiency with respect to the one presented in [45] is achieved, while reducing the nonprompt lepton contamination by about a factor of two to four, depending on the lepton p_T (see Fig. 5.3). The contribution of the nonprompt lepton background is negligible in the four-lepton final state, therefore the selection criteria on Lepton MVA in four-lepton final state is loosened with respect to the one used in three-lepton analysis. Other selection criteria are quite similar to the ones presented in the two previous analyses and are summarized in Table 8.1. The loose identification criterion is additionally tuned for the nonprompt lepton background estimation, which will be discussed in Section 8.3.

Electrons		
Criterion	Loose	Tight
$p_T > 10$ GeV	✓	✓
$\|\eta\| < 2.5$	✓	✓
$\|d_{xy}\| < 0.05$ cm	✓	✓
$\|d_z\| < 0.1$ cm	✓	✓
$SIP_{3D} < 8$	✓	✓
trigger emulation	✓	✓
0 or 1 missing pixel hits	✓	✓
closest b jet veto	✓	✓
$I_{mini} < 0.4$	✓	✓
Lepton MVA $<$ 0.4 (-0.4)	-	✓

Muons		
Criterion	Loose	Tight
$p_T > 10$ GeV	✓	✓
$\|\eta\| < 2.4$	✓	✓
$\|d_{xy}\| < 0.05$ cm	✓	✓
$\|d_z\| < 0.1$ cm	✓	✓
$SIP_{3D} < 8$	✓	✓
closest b jet veto	✓	✓
$I_{mini} < 0.4$	✓	✓
Medium ID	✓	✓
Lepton MVA $<$ 0.4 (-0.4)	-	✓

Table 8.1: List of all quality criteria for electrons and muons for passing the loose and the tight selection in the $t\bar{t}Z$ analysis. The upper bound for the Lepton MVA in the brackets indicates the BDT discriminator value used for leptons in the four-lepton analysis.

Jets with $p_T > 30$ GeV and $|\eta| < 2.4$ are selected for the analysis. Jets are identified as originating from b quarks using the Deep CSV algorithm described in Section 5.2.4. Similar to previous analyses, jets within $\Delta R = 0.4$ of the leptons passing loose selection criteria are rejected.

8.2 STRATEGY

The previous analysis showed that the multilepton final state is quite pure in $t\bar{t}Z$, at least in the most significant bins which are characterized by high jet and b-tagged jet multiplicities. The main backgrounds originate from the SM processes with at least one top quark in the final states: tZq, tWZ, $t\bar{t}H$, etc. While the Lepton MVA discriminator helps in increasing the number of signal events, it also increases the number of events that originate from the processes producing 3 or more prompt leptons in the final state.

In the three-lepton final state we ask for three leptons to pass the lepton identification criteria and the p_T thresholds of 40, 20 and 10 GeV for leading, sub-leading and trailing leptons to ensure almost 100% efficiency in the event selection. On top of this we ask two leptons to form an OCSF pair with invariant mass 10 GeV close to a Z boson mass to suppress the contribution from the processes without a Z boson in the final state. Events are categorized according to N_j and N_b. For the $N_b \geq 1$ category, the $N_j = 2, 3, 4, > 4$ categorization is used and events with less than 2 jets are rejected. In the N_b

= 0 category, events are split according to $N_j = 1, 2, 3, > 3$ selection and events with no jets are rejected.

In the four lepton final state, the main background originates from ZZ production. We ask for 4 leptons to pass full identification criteria and exactly one OCSF pair with an invariant mass within 20 GeV of the Z boson in order to suppress the contribution from ZZ decaying into 4 leptons. To satisfy the high trigger efficiency requirement, the leading lepton p_T is asked to be greater than 40 GeV and $p_T > 10$ GeV for 3 remaining leptons. On top of all these requirements, we ask for the sum of all leptons charges to be equal to 0 and for at least 2 jets. Two signal regions are defined with the $N_b = 0$ and $N_b > 0$ selection, which are added to the previously mentioned 12 categories in the three-lepton final state for the $t\bar{t}Z$ inclusive cross section measurement.

The summary of all signal regions used in the $t\bar{t}Z$ inclusive cross section measurement is shown in Table 8.2.

$\mathbf{N}(\ell)$	N_j	N_b
3	1, 2, 3, ≥ 4	0
3	2, 3, 4, ≥ 5	1, ≥ 2
4	≥ 2	0, ≥ 1

Table 8.2: The signal regions categorized by number of leptons, jets and b-tagged jets used in the $t\bar{t}Z$ inclusive cross section measurement.

8.3 BACKGROUND ESTIMATION

Several sources of backgrounds contribute to the three-lepton and four-lepton final states targeted in this analysis. These backgrounds can be subdivided into two categories: processes that can pass the selection with only prompt leptons and processes which contribute to a selection with nonprompt leptons. The nonprompt lepton background is largely eliminated by the lepton MVA selection applied to leptons, but remains significant nonetheless. Similar to the previous two analyses, the tight-to-loose technique is used for the nonprompt lepton background estimation.

Several processes lead to a final state with three or more prompt leptons and several jets. The largest contributions come from diboson production processes, namely WZ and ZZ. These processes tend to have a small jet multiplicity originating from ISR or FSR, so they are significantly reduced in the signal categories of this analysis, which require at least 2 jets in $N_b \geq 1$ categories, and they contribute mostly in low N_b categories. All these processes are validated in dedicated control regions.

Another significant source of background events originates from the processes with at least one top quark produced in association with the W, Z, or BEH bosons or off-shell photon, i.e. $t\bar{t}H$, tWZ, tZq, tHq, tHW, ttVV, and $t\bar{t}t\bar{t}$. Processes that involve a photon ($Z\gamma$ and $tt\gamma$) are grouped in the $X\gamma$ category and are validated in data events, where the invariant mass of the three leptons is close to a Z boson mass. The last category is the remaining rare SM processes, that have not been measured at all. The main contribution to this category originates from the three-boson processes. For all aforementioned categories, the number of events in signal regions are estimated from MC simulation scaled to the integrated luminosity of 77.5 fb^{-1}.

8.3.1 *Nonprompt lepton background*

The contribution from the nonprompt lepton background is estimated with the tight-to-loose technique described in Subsection 6.4.1. The tight-to-loose ratio is measured in events that are fired by the auxiliary triggers listed in Table D.2. The event selection remains the same as discussed in Subsection 6.4.1. The electroweak scale factor used to subtract the contribution from the processes yielding a prompt lepton in the tight-to-loose ratio measurement region is calculated separately for each category in the p_{T}-η misidentification rate map. Examples of electroweak normalization control regions for electrons and muons in low-p_{T} barrel category are shown in Fig. 8.3.

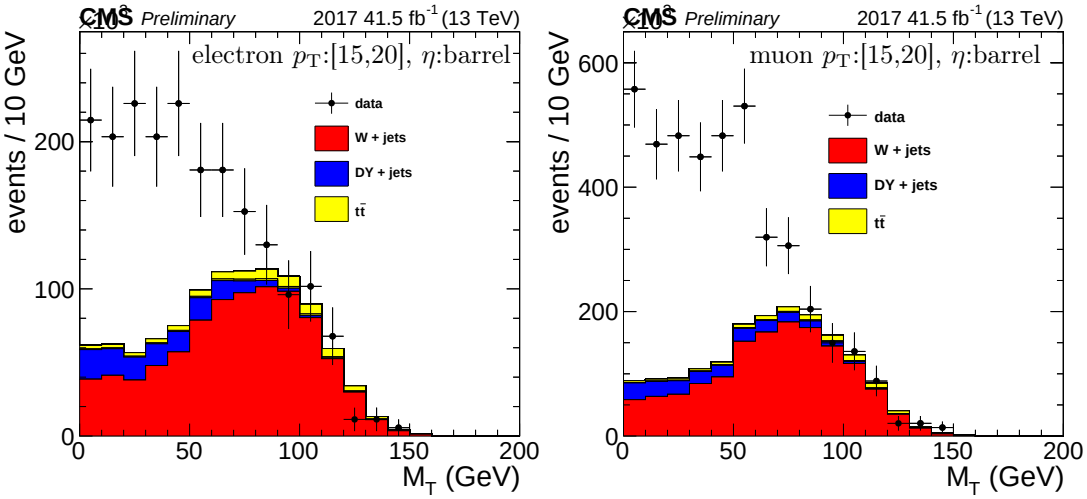


Figure 8.3: Distribution of M_{T} in the electroweak normalization control regions for electrons (left) and muons (right) in the barrel of the detector with lepton p_{T} from 15 to 20 GeV.

As discussed in Subsection 6.4.1, the probability for a nonprompt lepton to pass the selection criterion has a strong dependence on the p_{T} of the parton that gives the lepton. To mitigate this dependency we parametrize the misidentification rate as function of the parton p_{T}. To achieve this we use the following recipe: if the lepton passes tight selection requirements the p_{T} remains the same, while if it fails the tight selection, but passes the loose selection criterion, the p_{T} is corrected using the following formula:

$$p_{\mathrm{T}}^{corr} = 0.85 \times p_{\mathrm{T}}/p_{\mathrm{T}}^{\mathrm{ratio}},$$

where $p_{\mathrm{T}}^{\mathrm{ratio}}$ is defined in Subsection 5.2.3. The 0.85 factor is used to ensure continuity in the corrected p_{T} spectrum as a function of the lepton MVA. Throughout the presented analysis, the computation of all variables and the event selection is done using the modified lepton momentum p_{T}^{corr}, and p_{T} will henceforth refer to p_{T}^{corr}.

The loose object selection (see Table 8.1) is tuned to have a good match in the closure test both in $\mathrm{t\bar{t}}$ and DY MC simulations, separately for electrons and muons. First, the $p_{\mathrm{T}}^{\mathrm{ratio}}$ and closest jet Deep CSV values are tuned to have a good match for muons. With these variables, we also keep the misidentification rate at low level in order to have a good statistical precision for the nonprompt lepton background estimation. In the next step, we tune the electron MVA values to have a good match in the closure test for electrons in the barrel and the endcap. The measured misidentification rate in data for electrons and muons are shown in Fig. 8.4.

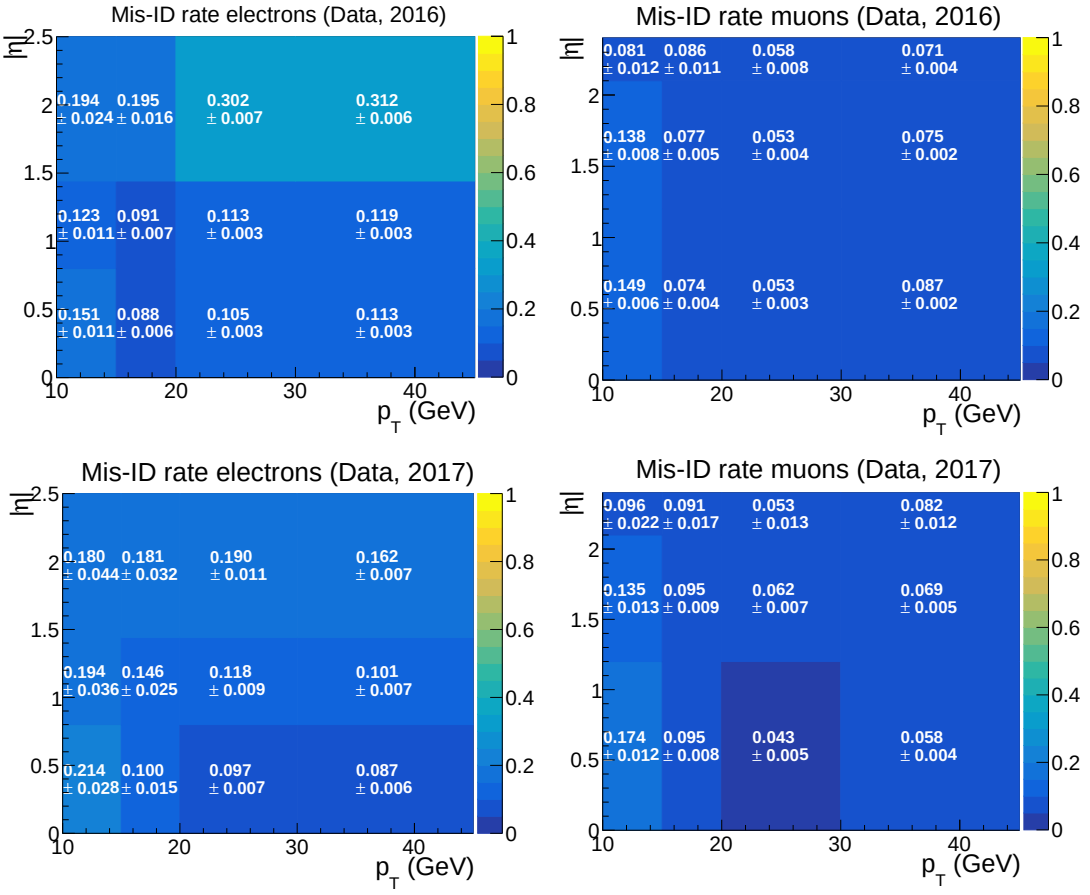


Figure 8.4: Misidentification rates measured in data as a function of cone-corrected p_{T} and absolute value of η after MC electroweak processes subtraction for electrons (left) and muons (right) in 2016 (top) and 2017 (bottom) data sets.

The method described above to predict backgrounds with at least one nonprompt lepton is validated using simulation and data events in control regions. The results of this validation can be seen in Fig. D.1 and D.2. Events with three leptons without a Z boson candidate provide a suitable data control sample for the nonprompt leptons stemming from the $\mathrm{t\bar{t}}$ process. Fig. 8.5 shows the predicted and observed yields for different lepton flavors as a function of the trailing lepton p_{T}, N_{j} and N_{b}. In general, a good agreement between predicted and observed yields is obtained, apart from the tails of N_{j} and N_{b} distributions where the nonprompt lepton background does not give a dominant contribution. Fig. 8.6 depicts the predicted and observed yields in three-lepton with a Z boson candidate control region enriched in nonprompt lepton background originating from the DY process. To ensure the orthogonality to the signal region, we also require 0 b jets in the selection, N_{j} to be equal 0 or 1 and low p_{T}^{miss} and M_{T} for the third lepton not used in forming the Z boson candidate. Good agreement is observed for the trailing lepton in the low-p_{T} category and overall in all lepton flavors. Based on these studies a systematic uncertainty of 30% is assigned, while the statistical uncertainty on the prediction of this background ranges from 5 to 50%, depending on the measurement bin.

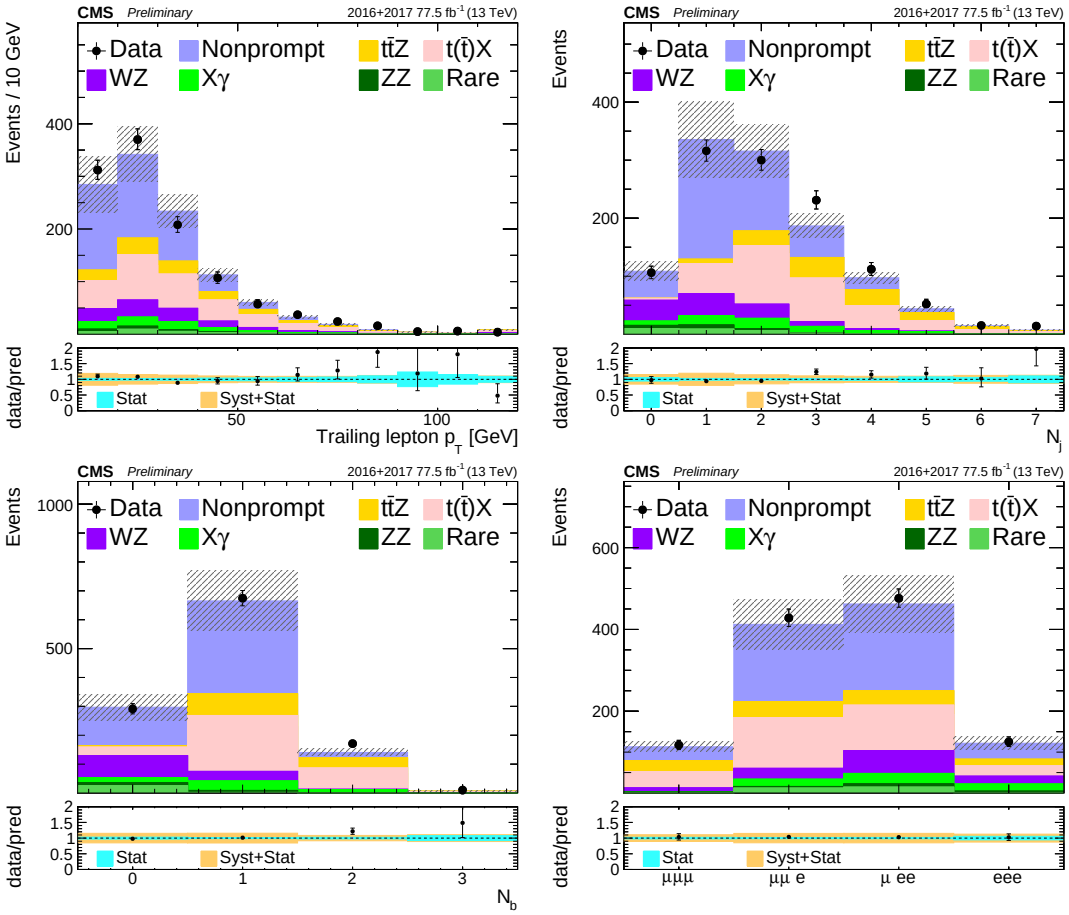


Figure 8.5: Predicted and observed yields as a function of trailing lepton p_{T}, N_{j}, N_{b} and the flavor composition in a region enriched in nonprompt lepton background originating from the $\mathrm{t\bar{t}}$ process. The contribution from this type of background is estimated using the tight-to-loose technique described in Section 6.4.1. The lower panels show the ratio of the data to the theoretical predictions. The inner band gives the statistical uncertainty in the ratio, and the outer band the total uncertainty.

8.3.2 *WZ and ZZ production*

Diboson production is one of the largest backgrounds in the signal region with low N_{j} and N_{b}. For WZ and ZZ processes the overall production rate, as well as several relevant kinematic distributions predicted through simulation, can be tested in respective data samples that are orthogonal to the signal selection. Events with three leptons, two of which form an OCSF pair 10 GeV close to a Z boson mass, and no b-tagged jets are used to validate the WZ background prediction. Four-lepton events with two Z boson candidates are used to constrain the uncertainties in the prediction of the ZZ yield. Fig. 8.7 presents the observed and predicted event yields in these control regions. Individual normalization uncertainty of 10% is assigned to WZ and ZZ processes, while for events with $N_{\mathrm{j}} \geq 3$ an additional 20% uncertainty is assigned to the WZ background prediction. We furthermore investigate potential mismodeling effects associated with the WZ production with heavy-flavor quark pairs originating from gluon splitting in a data sample enriched with DY events. In a control region with a Z boson candidate and two b-tagged jets, a comparison of the distributions of the relative angle between the two b jets in data and simulation is used to assess the uncertainty due to possible

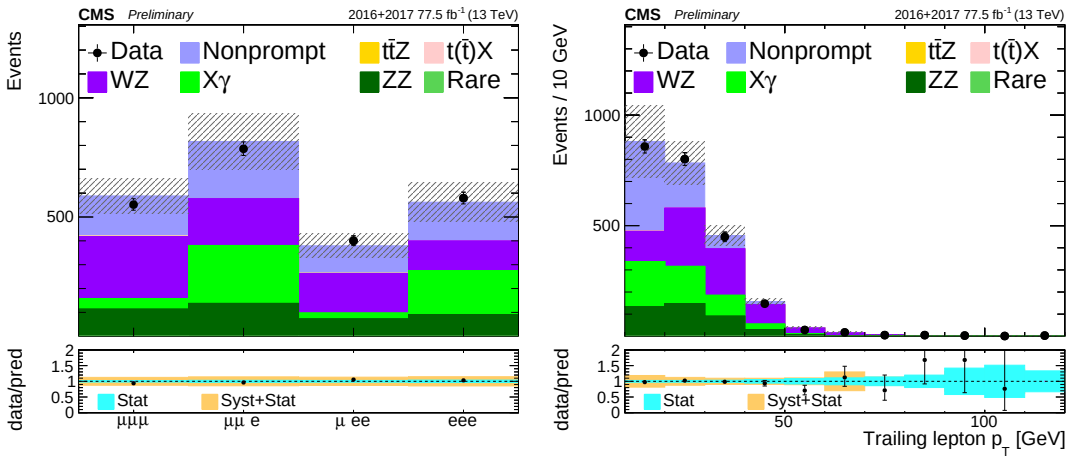


Figure 8.6: Predicted and observed yields from the trailing lepton p_{T} and the flavor composition in a region enriched in nonprompt lepton background originating from the DY process. The contribution from this type of background is estimated using the tight-to-loose technique described in Section 6.4.1. The lower panels show the ratio of the data to the theoretical predictions. The inner band gives the statistical uncertainty in the ratio, and the outer band the total uncertainty.

mismodelling. It is estimated to be less than 20%. Propagating this uncertainty to the rate of simulated WZ events with gluon splitting, the uncertainty on the rate of WZ events with $N_{\mathrm{b}} \geq 1$ is estimated to be 8%.

8.3.3 *Background caused by the internal conversions*

Events produced in Zγ process form a copious source of three lepton events through asymmetric conversions of the photon. In the case of leptonically decaying Z bosons, most Zγ events occur through FSR of a photon by one of the leptons. This photon can then convert asymmetrically, leading to a total of three leptons, with a mass compatible with the Z boson. A Zγ enriched data sample is obtained by selecting three leptons that pass the same selection as the three-lepton final state and requiring the mass of the three-lepton system to be compatible with the Z mass. From the overall good agreement in all kinematical variables shown in Fig 8.8, the 20% uncertainty on the normalization is assigned to the all processes in the Xγ category. The contribution from all the processes in Xγ category to the signal regions are estimated using MC simulation, which is performed with the MADGRAPH5_aMC@NLO generator.

8.3.4 *Other processes*

The most significant processes among the remaining background processes in the high N_{j} and N_{b} categories are the ones that produce one top quark and a Z boson in the final state, namely tZq and tWZ. These and other processes giving at least one top quark in the final state are grouped in the t$\bar{\mathrm{t}}$X category (for the full list of the processes entering this category see Table D.1). The contribution in last category - rare SM backgrounds - originates from three-boson production of W and Z bosons in any combination. The contributions from t$\bar{\mathrm{t}}$X and the rare SM backgrounds category are estimated with the produced yields from MC simulation multiplied by the appropriate scale factors.

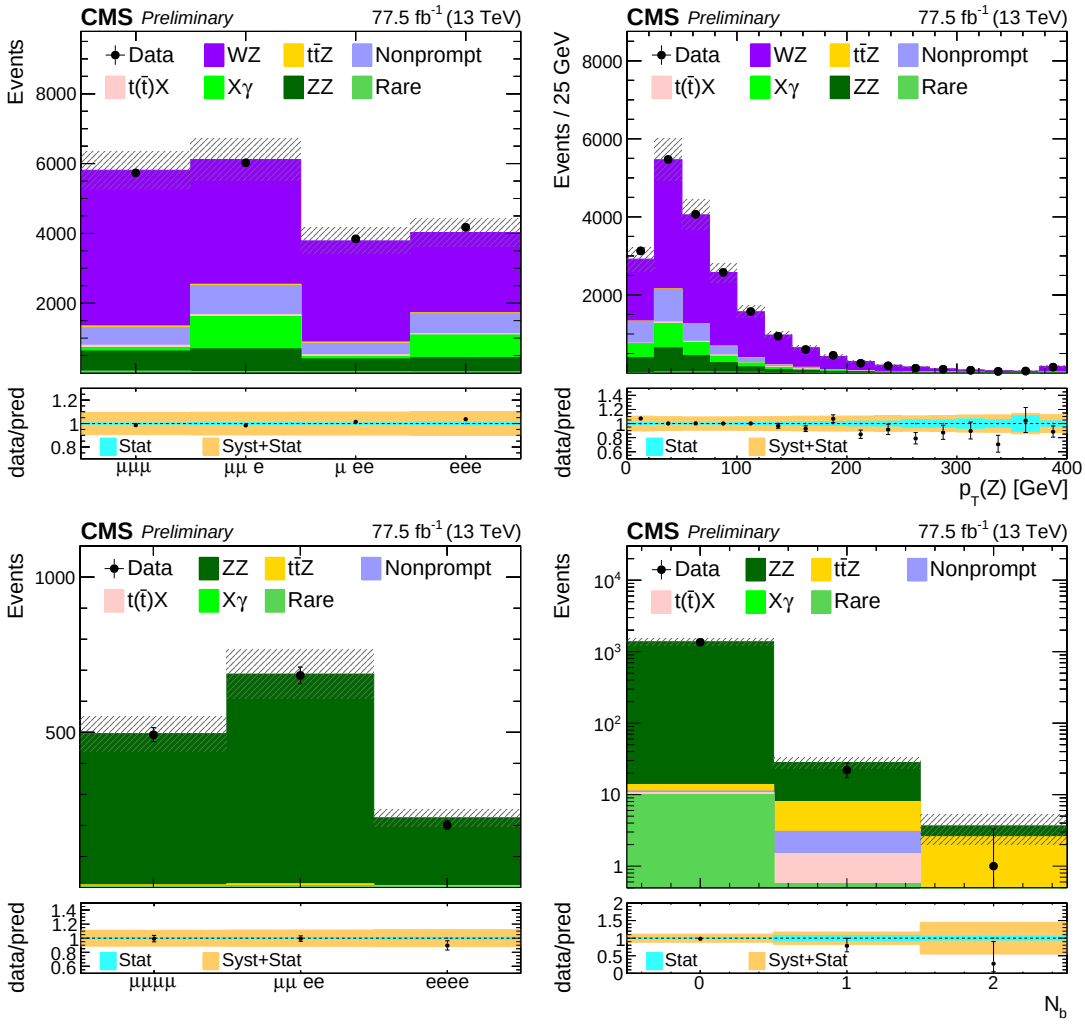


Figure 8.7: Distributions of the predicted and observed yields versus lepton flavor (upper left), and the reconstructed transverse momentum ($p_{\mathrm{T}}(\mathrm{Z})$) of the Z boson candidates (upper right) in the WZ-enriched data control region, and versus lepton flavor (lower left) and N_{b} (lower right) in the ZZ-enriched control region. The shaded band represents the total uncertainty in the prediction of the signal and background. The lower panels show the ratio of the data to the estimated predictions. The inner band gives the statistical uncertainty in the ratio, and the outer band the total uncertainty.

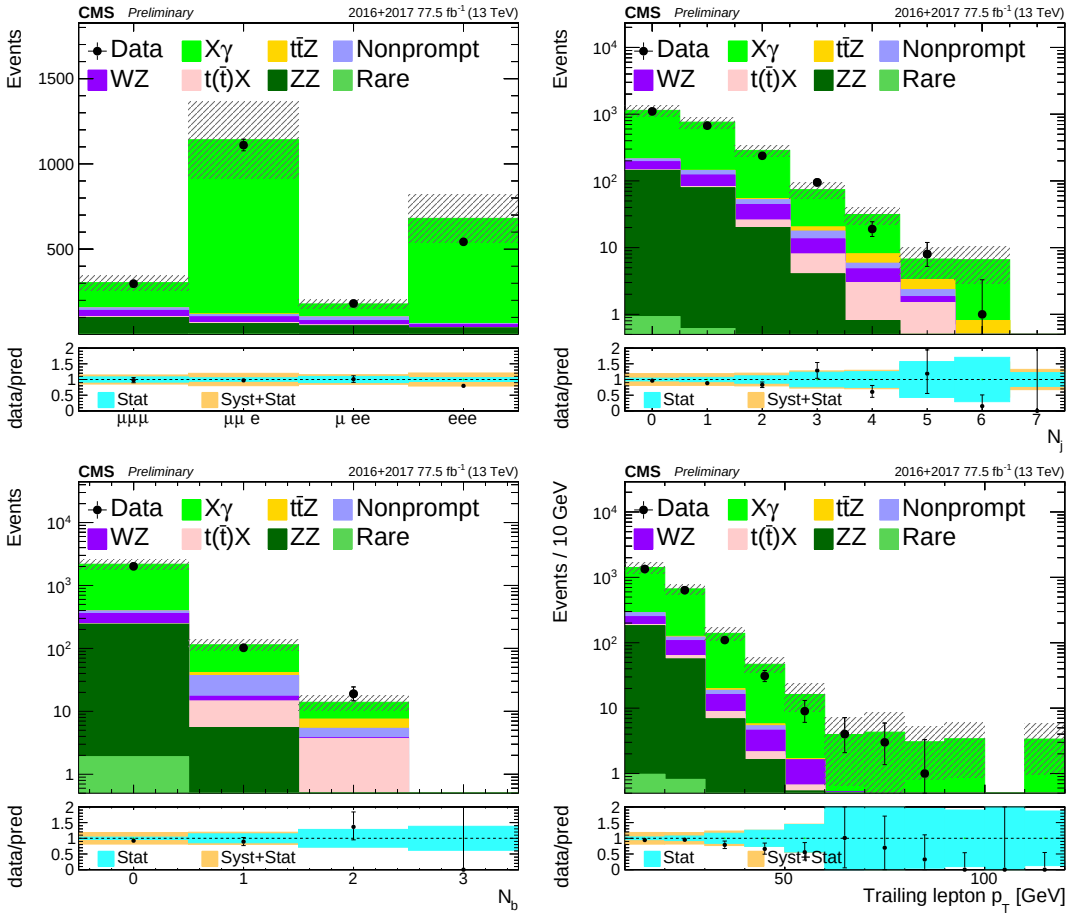


Figure 8.8: Distributions of the predicted and observed yields versus lepton flavor, N_{j}, N_{b} and trailing lepton p_{T} in the Zγ-enriched control region. The shaded band represents the total uncertainty in the prediction of the signal and background. The lower panels show the ratio of the data to the estimated predictions. The inner band gives the statistical uncertainty in the ratio, and the outer band the total uncertainty.

8.4 SYSTEMATIC UNCERTAINTIES

The systematic uncertainties affecting the signal selection efficiency and background yields are summarized in Table 8.3. The table indicates both the range of variations observed in the different bins of the analysis and their impact on the final result. The table also indicates whether the uncertainties are treated fully correlated or completely uncorrelated between the 2016 and 2017 data sets.

Source	Uncertainty range (%)	Correlated in 2016 and 2017	Impact on the $t\bar{t}Z$ cross section (%)
Integrated luminosity	2.5	×	2
PU modeling	1–2	✓	1
Trigger	2	×	2
Lepton ID efficiency	4.5–6	✓	4
JEC	1–9	✓	2
JER	0–1	✓	1
B tagging light flavor	0–4	×	1
B tagging heavy flavor	1–4	×	2
Choice in μ_R and μ_F	1–4	✓	1
PDF choice	1–2	✓	1
Color reconnection	1.5	✓	< 1
Parton shower	1–8	✓	1
WZ cross section	10–20	✓	3
WZ + heavy flavor	8	✓	1
ZZ cross section	10	✓	1
$t\bar{t}X$ background	10–15	✓	3
$X\gamma$ background	20	✓	1
Nonprompt lepton background	30	✓	< 1
Rare SM background	50	✓	2
Stat. unc. in nonprompt bg.	5–50	×	< 1
Stat. unc. in rare SM bg.	5–100	×	< 1
Total uncertainty			7

Table 8.3: Summary of the sources, magnitudes, treatments, and effects of the systematic uncertainties in the $t\bar{t}Z$ inclusive cross section measurement. The first column indicates the source of the uncertainties, the second column shows the corresponding input uncertainty range on each background source and the signal. The third column indicates how correlations are treated between uncertainties in the 2016 and the 2017 data, and the fourth column shows the resulting uncertainty in the $t\bar{t}Z$ cross section.

The procedure for estimating and propagating the uncertainties related to the luminosity and PU reweighting does not change with respect to the one described in Chapter 6.5. The trigger selection efficiencies in the simulation are found to be in good agreement with the one measured in data. The uncertainties related to the trigger efficiency discussed in Section 8.1 is propagated to the signal region and results in approximately 2% uncertainty.

Lepton selection efficiencies are measured using a tag-and-probe method described in Section 5.3 in bins of lepton p_T and η. The differences between the two measurements in data and in simulation range from 2-7% for $10 < p_T < 20$, 0.5% for the intermediate region (20-50 GeV) and raises up to 1-2% for $p_T > 50$ GeV. These differences are used to scale the yields obtained in the simulation. The systematic uncertainties related to this source vary between 4.5–6% on the signal and background yields.

The validity of the lepton scale factors, applied on simulation to bring the lepton selection efficiency in agreement with the one in data, is validated in events with high hadronic activity originating from the $t\bar{t}$ process. Events containing exactly one electron and one muon, at least 2 jets and 1 b-tagged jet are selected with a single lepton trigger for this study. One of the leptons is used as a tag lepton in the event, while for another the tight selection efficiency is measured. This control sample has some small contamination from nonprompt leptons which need to be subtracted out. We use the data events in the same selection region with one difference, the leptons being same-charge instead of opposite charge, to estimate this nonprompt lepton contamination. This subtraction relies on the fact that there are equal amounts of same-charge and opposite-charge lepton pairs with one prompt and one nonprompt lepton. This assumption is correct to a good extent. To take into account the residual difference we use simulation to correct for it. The uncertainty (30%) on this subtracted background is propagated to the efficiency calculation. The measured lepton selection efficiency in this data control region is compared to the one obtained in the simulation. This comparison is shown in Fig. 8.9. The efficiencies obtained in simulation are already corrected using the lepton scale factors which we want to validate here. We observe very good agreement between data and simulation efficiencies. Any discrepancy seen at the lowest p_T bins is covered by the uncertainties. Unfortunately a similar validation could not be made for electron selection scale factors as the nonprompt electron background contamination in such a control sample is found to be too large to give conclusive results.

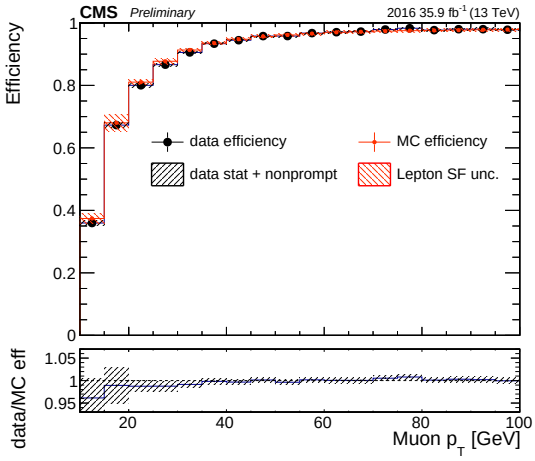


Figure 8.9: Efficiency to pass the tight selection criteria for the probed muon measured using in control region, enriched in the $t\bar{t}$ events in data (black) and MC simulation (red).

The uncertainties related to imperfect knowledge of JEC and on the correction of the b tagging efficiencies are estimated in the same way as described in Section 6.5. In addition to these uncertainties, the uncertainty on JER is also propagated to the yields of the signal and background processes in the analysis.

The procedures for estimating and propagating the theoretical uncertainties due to the choice of the renormalization and factorization scales, and PDF remain similar as

described in Section 6.5. In the parton shower simulation, the uncertainty from the choice of μ_F is estimated by varying the scale of ISR and FSR up by factors of 2 and $\sqrt{2}$ and down by factors of 0.5 and $1/\sqrt{2}$, respectively, as suggested in Ref. [131]. The default setup in PYTHIA includes a multiple-parton-interaction-based (MPI) model of color reconnection with early resonance decays switched off. To estimate the uncertainty from this choice of model, the analysis is repeated with three other color reconnection models within PYTHIA: the MPI-based scheme with early resonance decays switched on, a gluon-move scheme [132], and a QCD-inspired scheme [133]. The total uncertainty from color reconnection modeling is estimated by taking the maximum deviation from the nominal result and amounts to 1.5%. The magnitude and shape of all uncertainties for the $t\bar{t}Z$ process are shown in Fig 8.10, while for the background WZ and $t\bar{t}X$ processes they are shown in Appendix. D.4.

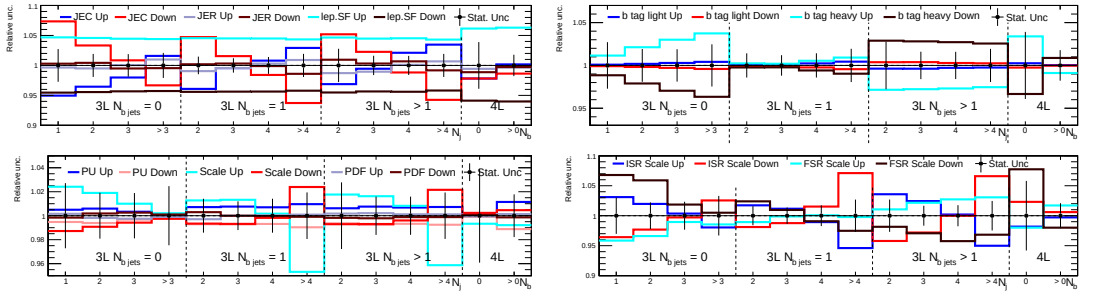


Figure 8.10: Input systematic uncertainties for $t\bar{t}Z$ in each of 14 signal regions used for signal strength modifier extraction. Uncertainties for JEC, JER, lepton scale factors, heavy and light flavor b tagging scale factors, PU, μ_R and μ_F scales, PDF and modeling of the parton shower are shown.

The theoretical uncertainties for the processes in the $t\bar{t}X$ category similarly to the $t\bar{t}W$ cross section measurement are taken from [129], apart from the uncertainty on the recently discovered tZq process [3]. The measured cross section for the tZq process reached a 15% uncertainty level, which is taken as the uncertainty on the normalization in this analysis. The overall uncertainty in the normalization of the rare SM processes background category is chosen to be 50% [134].

8.5 RESULTS

The signal extraction is performed using the events with three and four leptons and various jet and b jet categories described in Section 8.2. The observed data, as well as the predicted signal and background yields are shown in Fig. 8.11. The measured signal strength modifier is estimated by minimizing the likelihood function from Eq. 6.4 and is found to be $1.29\ ^{+0.08}_{-0.07}(\text{stat.})\ ^{+0.08}_{-0.08}(\text{syst.})$. The signal strengths separately for the three- and four-lepton channels and in different three-lepton flavor categories are shown in Tables 8.4 and 8.5, respectively. The largest deviation in the signal strength modifier was observed in the three-electron channel, where statistical uncertainty significantly larger with respect to the other flavor channels was observed. If this high signal strength was due to an underestimated signal electron efficiency in simulation, an excess should appear in three-electron or four-electron channels in WZ and ZZ control regions presented in Fig. 8.7. However, no such excess was observed, therefore, we attribute this excess to a statistical fluctuation in data.

Fit configuration	Measured signal strength modifier
3ℓ	$1.32\ ^{+0.08}_{-0.07}(\text{stat.})\ ^{+0.08}_{-0.08}(\text{syst.})$
4ℓ	$1.24\ ^{+0.20}_{-0.18}(\text{stat.})\ ^{+0.11}_{-0.09}(\text{syst.})$

Table 8.4: Measured $\text{t}\bar{\text{t}}\text{Z}$ signal strength modifiers for events with three and four leptons.

Flavor channel	Measured signal strength modifier
3 leptons	$1.32\ ^{+0.08}_{-0.07}(\text{stat.})\ ^{+0.08}_{-0.08}(\text{syst.})$
3μ	$1.17\ ^{+0.13}_{-0.12}(\text{stat.})\ ^{+0.10}_{-0.09}(\text{syst.})$
2μ1e	$1.38\ ^{+0.15}_{-0.15}(\text{stat.})\ ^{+0.11}_{-0.10}(\text{syst.})$
1μ2e	$1.32\ ^{+0.16}_{-0.15}(\text{stat.})\ ^{+0.10}_{-0.09}(\text{syst.})$
3e	$1.69\ ^{+0.21}_{-0.20}(\text{stat.})\ ^{+0.13}_{-0.11}(\text{syst.})$

Table 8.5: Measured $\text{t}\bar{\text{t}}\text{Z}$ signal strength modifiers in different combination of three-lepton flavor categories.

The impact of each systematic uncertainty can be seen in Fig. D.5, D.6 and D.7. The impact is evaluated in the following way. For a nuisance parameter under consideration, the post-fit value is taken with upward and downward one standard deviation; other systematic uncertainties are treated as usual in the maximization of the likelihood function. With this configuration, the signal strength modifier is reevaluated and the difference between obtained value and initial signal strength modifier is considered as the impact of the nuisance parameter. The total uncertainty of the $\text{t}\bar{\text{t}}\text{Z}$ signal strength modifier is found to be 8%. The statistical and systematic uncertainties contribute approximately equally to the total uncertainty. Among the systematic uncertainties the ones that give the largest contributions are those related to lepton selection efficiency modeling, and the normalization uncertainties on WZ and $\text{t}\bar{\text{t}}\text{X}$ backgrounds (see Fig. D.7).

In order to extract the cross section of the $\text{t}\bar{\text{t}}\text{Z}$ production, the cross section value obtained from the ME of the sample simulated with the MADGRAPH5_aMC@NLO generator with applied $m_{\ell\ell} > 70$ GeV selection for the OCSF pair is divided by the branching ratio of a Z boson decaying to charged and neutral leptons, $\text{BR}(Z \to \ell\ell, \nu\nu) = 0.301$ [10] and is multiplied by the signal strength modifier. The contribution from $\gamma^* \to \ell\ell$ is excluded from this estimation. The measurements in each data set come out to be consistent within their statistical precision as expected:

$$2016 : \sigma(pp \to \text{t}\bar{\text{t}}\text{Z}) = 0.89\ ^{+0.07}_{-0.07}(\text{stat.})\ ^{+0.08}_{-0.07}(\text{syst.})\text{pb},$$

$$2017 : \sigma(pp \to \text{t}\bar{\text{t}}\text{Z}) = 1.02\ ^{+0.07}_{-0.07}(\text{stat.})\ ^{+0.06}_{-0.06}(\text{syst.})\text{pb}.$$

The cross section measured in the combined 2016 and 2017 data sets is:

$$\sigma(pp \to \text{t}\bar{\text{t}}\text{Z}) = 0.95\ ^{+0.05}_{-0.05}(\text{stat.})\ ^{+0.06}_{-0.05}(\text{syst.})\text{pb}.$$

With respect to the earlier measurements [45, 135], the total uncertainty on the $\text{t}\bar{\text{t}}\text{Z}$ inclusive cross section is reduced by about 40%. While this improvement is primarily due to a better lepton selection procedure and the detailed studies of its performance in simulation, improved estimation of the trigger and b tagging selection efficiencies in

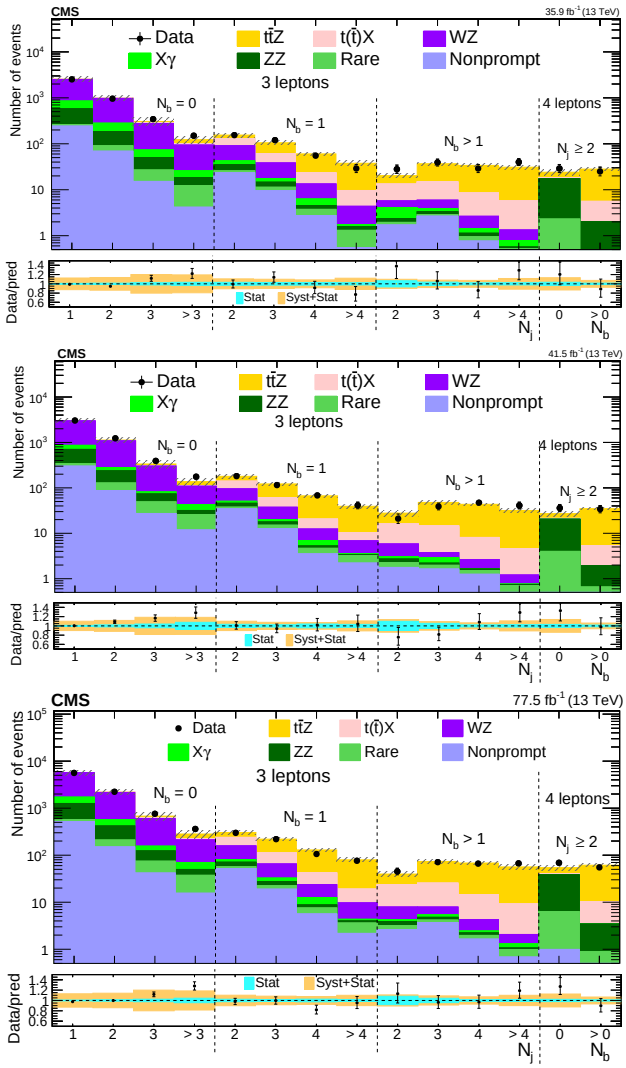


Figure 8.11: Observed data as a function of N_j and N_b for events with three and four leptons, compared to the simulated signal and background yields, as obtained from the fit in 2016 (top), 2017 (middle) and combined (bottom) data sets. The hatched band shows the total uncertainty associated with the signal and background predictions, as obtained from the fit.

simulation, and the reduced statistical uncertainty through the addition of more data also play important role. The reported result is the first experimental measurement that is more precise than the state-of-the-art theoretical calculations estimated at NLO QCD and electroweak accuracy for the $t\bar{t}Z$ production [129].

A $t\bar{t}Z$ signal-enriched region is defined by selecting events with $N_b \geq 1$ and $N_j \geq 3$ (2) for the three (four)-lepton channels. In this region, the signal purity is about 65% and the processes contaminating the most are the WZ and $t\bar{t}X$ processes, as it can be seen in Table 8.6; also, in the electron channels, the substantial contribution originates from the processes yielding at least one nonprompt lepton. In Fig. 8.12 several kinematical variable distributions in the $t\bar{t}Z$ enriched region are shown; same distributions, separately for 2016 and 2017 data sets, are shown in Fig. D.8 and D.9. Overall, good agreement is observed both in variables related to the hadronic activity (N_j and N_b), and to the

leptons ($m_{\ell\ell}$, $p_T(Z)$ and $\cos\theta^{*}_{Z}$[2]). Using the $t\bar{t}Z$ inclusive cross section measurement described above as a basis, for the first time differential cross sections as functions of $p_T(Z)$ and the $\cos\theta^{*}_{Z}$ were calculated. These measurements are presented in the Fig. 8.13. In general, the absolute differential cross section is measured to be slightly higher than the prediction, while the shape is described well.

Process	$(\mu)\mu\mu\mu$	$(\mu)\mu\mu$ e	$(\mu/e)\mu$ ee	(e)eee	Total
$t\bar{t}Z$	143.4 ± 7.1	122.1 ± 6.1	111.8 ± 5.5	77.4 ± 3.9	454.7 ± 22.1
$t\bar{t}H$	4.1 ± 0.5	3.5 ± 0.4	3.3 ± 0.4	2.1 ± 0.3	13.1 ± 1.6
$t\bar{t}X$	34.2 ± 4.2	28.2 ± 3.4	23.9 ± 2.9	18.4 ± 2.3	104.6 ± 12.7
WZ	17.5 ± 4.7	15.1 ± 4.2	10.2 ± 2.8	11.2 ± 3.1	53.9 ± 14.5
$X\gamma$	1.8 ± 1.8	2.1 ± 2.7	0.6 ± 0.6	4.6 ± 1.6	9.0 ± 3.9
ZZ	2.8 ± 0.4	2.7 ± 0.4	2.5 ± 0.3	2.2 ± 0.3	10.1 ± 1.3
Rare	2.9 ± 1.5	2.1 ± 1.1	1.8 ± 1.0	1.4 ± 0.7	8.3 ± 4.2
Nonprompt	6.9 ± 2.9	10.7 ± 4.0	6.9 ± 2.9	8.5 ± 3.5	33.1 ± 12.6
Total	213.5 ± 12.3	186.5 ± 11.5	161.0 ± 9.0	125.7 ± 8.2	686.8 ± 39.6
Observed	192	175	152	141	660

Table 8.6: Predicted and observed yields with total uncertainties in the signal-enriched sample of events. The yields are presented for the signal and the background processes in each flavor category in the three- and four-lepton channels.

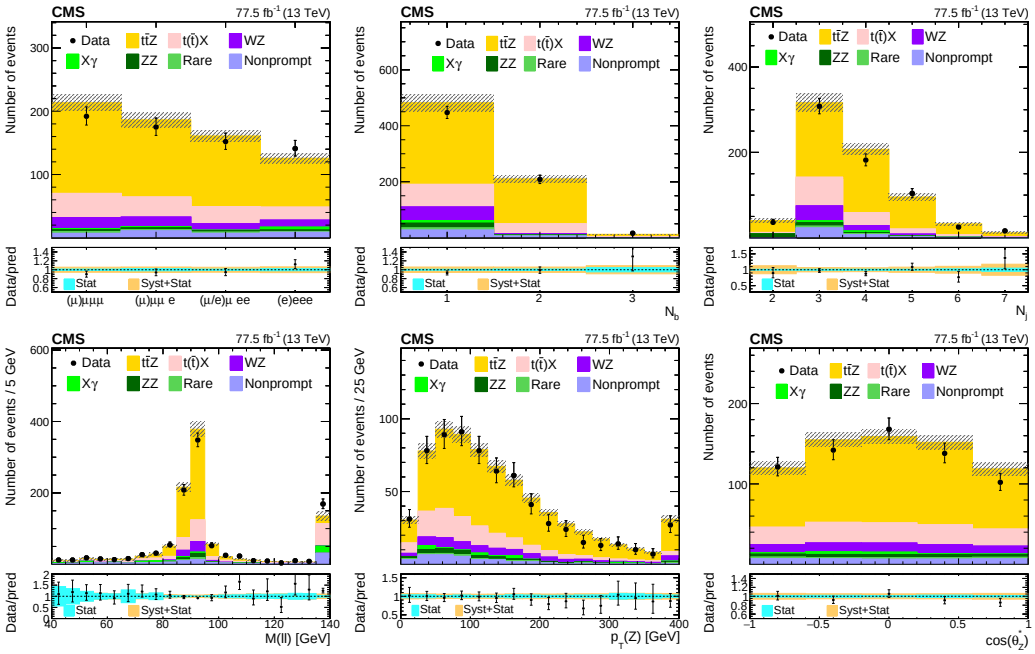


Figure 8.12: Observed data in a $t\bar{t}Z$ dominated region, compared to signal and background yields, as obtained from the fit. Event distributions are shown as a function of the lepton flavor (upper left), N_b (upper middle), N_j (upper right), dilepton invariant mass of OCSF pair (lower left), $p_T(Z)$ (lower middle), and $\cos\theta^{*}_{Z}$ (lower right). The hatched band shows the total uncertainty associated with the signal and background predictions, as obtained from the fit.

2 cosine of the angle between the Z boson and the negatively charged lepton from the Z boson as measured in the Z rest frame

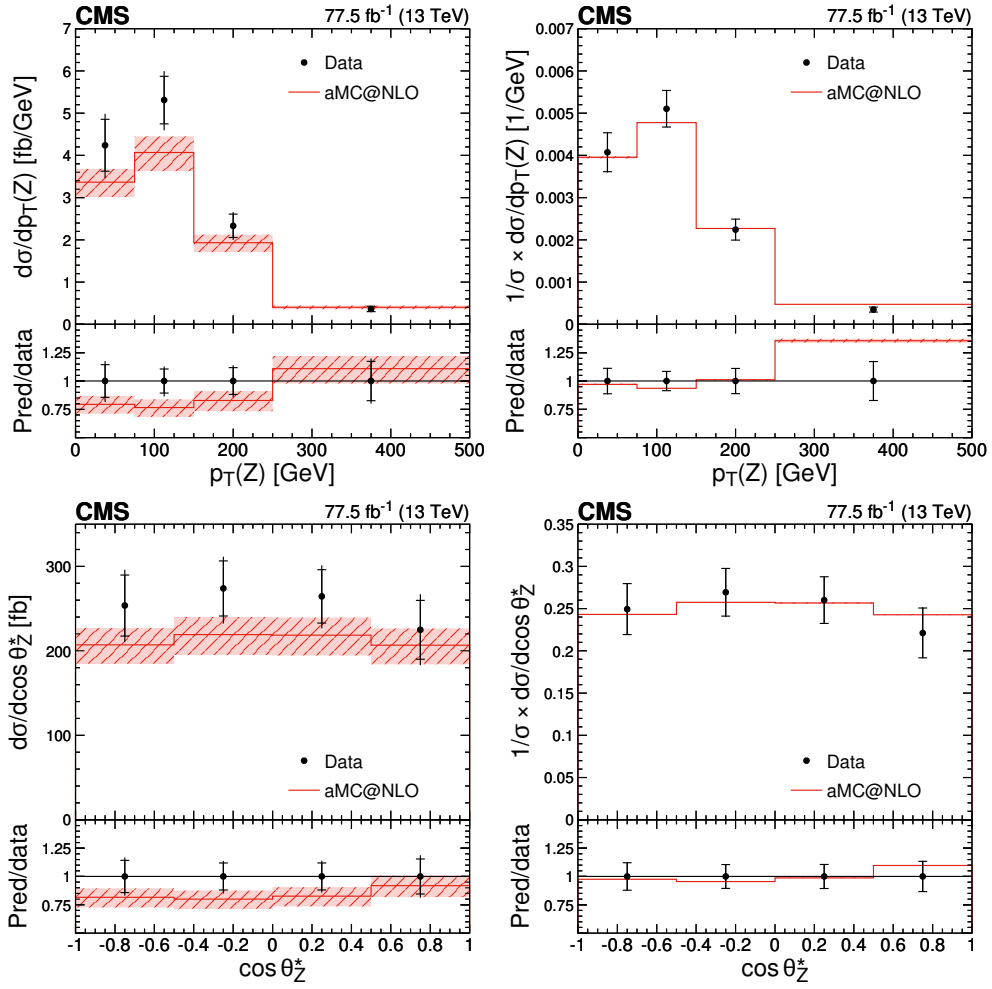


Figure 8.13: Measured differential $t\bar{t}Z$ production cross sections in the full phase space as a function of the $p_T(Z)$ (top row), and $\cos\theta^*_Z$, as defined in the text (bottom row). Shown are the absolute (left) and normalized (right) cross sections. The data are represented by the points. The inner (outer) vertical bars indicate the statistical (total) uncertainties, respectively. The histogram shows the prediction from the MADGRAPH5_aMC@NLO MC simulation, and the hatched band indicates the theory uncertainties in the prediction. The lower panels display the ratio between prediction and measurement. The figures are made by Joscha Knolle.

9

CONCLUSIONS AND OUTLOOK

The first part of this thesis is dedicated to the searches for strong SUSY production in collisions events with at least 3 electrons or muons, jets, and missing transverse momentum. The presented analysis follows an inclusive approach to achieve sensitivity for a wide range of BSM physics processes that produce three or more prompt leptons. Simplified SUSY models, featuring gluino and 3rd generation squark pair production are used as benchmark processes in order to optimize the event selection and to interpret the results in terms of exclusion limits on the masses of the SUSY particles. To maximize the sensitivity for discovery, events in the multi-lepton category are categorized according to the presence of a Z boson candidate, b quarks in the final state, high and low hadronic activity, missing transverse momentum and the transverse mass reconstructed with a lepton and the missing transverse momentum vector.

The main backgrounds in considered signal regions originate either from the processes that produce at least one nonprompt lepton that passes all identification and isolation requirements or from processes yielding at least three prompt leptons. The nonprompt lepton background is estimated with a tight-to-loose technique, which is validated both in the MC simulation and in a dedicated data control region enriched in nonprompt lepton background. The WZ and $\mathrm{t\bar{t}Z}$ processes are the main backgrounds in the regions with an on-shell Z boson candidate. The contribution from these processes is estimated in MADGRAPH5_aMC@NLO MC simulation and the normalization is obtained from a simultaneous fit using two control regions enriched in one of the processes.

For the analyzed data set of 35.9 fb^{-1} collected by the CMS experiment in 2016, no significant excess in the considered signal regions was observed except a few regions at high leading lepton p_{T} and hadronic activity. No accumulation of the events in any other kinematical variable distributions has been observed, therefore these events are considered as a statistical fluctuation. Exclusion limits at 95% confidence level are computed for four different SUSY simplified models. In the models featuring gluino pair production, namely T1tttt and T5qqqqWZ, gluinos with masses smaller than 1610 GeV are excluded for a massless neutralino. In the T6ttWW, the bottom squark masses are excluded up to 840 GeV for charginos lighter than 200 GeV. For T6ttHZ model, the $\widetilde{\mathrm{t}}_2$ mass is excluded up to 720, 780, or 710 GeV for models with an exclusive $\widetilde{\mathrm{t}}_2 \to \widetilde{\mathrm{t}}_1\mathrm{H}$ decay, an exclusive $\widetilde{\mathrm{t}}_2 \to \widetilde{\mathrm{t}}_1\mathrm{Z}$ decay, or an equally probable mix of those two decays, respectively.

The data volume of the proton collisions collected at the LHC in the whole Run II (approx. 140 fb^{-1}) is approximately 4 times larger than in the analysis presented in this PhD thesis. The increased data volume and usage of multilepton triggers with lower p_{T} thresholds and additional requirement either on missing transverse momentum or hadronic activity can improve the sensitivity of the analysis for the SUSY models in the compressed region. An alternative "in situ" technique to measure the lepton misidentification rate in the $\mathrm{t\bar{t}}$ events can be used. With this technique, the misidentification rate

can be measured directly in the process that contributes most significantly in the signal regions. Thus, the dependency on the flavor composition of the jets can be avoided which would lead to the reduction of the systematic uncertainty for the main background category in the most off-Z signal regions. Moreover, the classic approach which uses numerous signal regions for event categorization can be substituted with parametrized machine learning techniques, in which additionally to the traditional set of event-level kinematic variables a few parameters describing the SUSY model scenario are added [136].

The second part of the thesis is dedicated to the inclusive cross section measurement for $t\bar{t}W$ and $t\bar{t}Z$ processes. The measurement of the $t\bar{t}W$ production is performed in the same-charge dilepton final state and reached a precision level of 20% with the data collected by the CMS experiment in 2016 (35.9 fb^{-1}). The cornerstone of this analysis is the usage of a multivariate analysis which exploits boosted decision tree classifier with gradient boosting to discriminate between the signal events and events originating from the main background process, namely $t\bar{t}$. The statistical and systematic uncertainties contribute approximately equally to the total uncertainty on the measured cross section. The statistical uncertainty can be reduced approximately by a factor 2 by using the data volume collected by the CMS experiment in the whole Run II, thereby the precision of the measurement would be determined by the size of the systematic uncertainty. To reduce the systematic uncertainty, the efforts should be concentrated on the reduction of uncertainties originating from the efficiency measurement of the trigger selection, the knowledge of the nonprompt lepton background and the normalization of the processes contributing to the $t\bar{t}X$ category. The nonprompt lepton background and its uncertainty can be reduced either by the usage of multivariate technique, such as used in the $t\bar{t}Z$ analysis or developing the "in-situ" tight-to-loose technique for greater precision estimation. Usage of a mix of single lepton and dilepton triggers will help to reach the full selection efficiency of the signal events thereby reducing uncertainty on the trigger efficiency selection. The dedicated multivariate technique to separate the $t\bar{t}W$ process from similar processes in the top quark sector, such as $t\bar{t}H$ and tHq, will be indispensable in the subsequent analysis. This will be the most challenging problem to address in the future $t\bar{t}W$ inclusive cross section measurement. In addition to the measurement of the $t\bar{t}W$ inclusive cross section in the same-charge dilepton final state with Run II data set, the next most prominent category - three-lepton final state - should be considered in the analysis. These two channels are the most important for subsequent measurements of the differential cross section and the $t\bar{t}W^+/t\bar{t}W^-$ charge asymmetry, which potentially can shed the light on the new physics in the top quark sector [47].

The data sample used to measure the $t\bar{t}Z$ production was collected by the CMS experiment during 2016 and 2017 (77.5 fb^{-1}). The inclusive cross section is measured in the three- and four-lepton final states reaching the precision level of 8%. The reported result is the first experimental measurement that is more precise than the state-of-the-art theoretical calculations estimated at NLO QCD and electroweak accuracy for the $t\bar{t}Z$ production. To reduce further the systematic uncertainties of the $t\bar{t}Z$ inclusive cross section measurement, the efforts should be concentrated on the reduction of the uncertainties related to the modeling of the lepton selection efficiency and the WZ and $t\bar{t}X$ background. The most significant contribution to the lepton scale factor (to match it to the one in data) uncertainty originates from the relatively low lepton p_T categories. Therefore, to suppress its contribution, the trailing lepton p_T requirement can be increased to 20 GeV in the next analysis. However, this increase will lead to approximately 15% reduction in the signal events. Given that the measurement is not anymore statistics limited, this increase in p_T can effectively improve the precision. Such

an increase will also automatically reduce the nonprompt lepton background which predominantly contribute with a low p_{T} lepton. To discriminate against the background contributing to the WZ and $\mathrm{t\bar{t}X}$ categories, multivariate techniques can be used similarly to the one used for the discovery of a single top production in association with a Z boson [3]. In the high N_{b} control region the purity of the $\mathrm{t\bar{t}Z}$ process reaches up to 80%. It would be very beneficial to perform the cross section measurements of $\mathrm{t\bar{t}Z}$ and tZq processes together and constrain the uncertainties of both processes simultaneously.

10

NEDERLANDSE SAMENVATTING

Het Standaard Model van de deeltjesfysica is één van de meest succesvolle theorieën uit de vorige eeuw. Het omvat onze kennis van alle gekende elementaire deeltjes, en vele van zijn voorspellingen werden experimenteel bevestigd in de voorbije decennia, met als hoogtepunt de ontdekking van het BEH boson in 2012 door de CMS en ATLAS Collaboraties. Hoewel het SM de huidige data zeer goed beschrijft, zijn er enkele fenomenen die nog niet in de theorie bevat zitten: de indicatie voor donkere materie in het universum, de assymetrie tussen materie en antimaterie en de neutrino oscilaties. Bovendien omvat het SM slechts drie van de vier fundamentele krachten: electromagnetische, zwakke en sterke interacties. Zwaartekracht, de dominante interactie op grote schaal, wordt niet beschreven door het SM.

Dankzij het grotere datavolume en de verhoogde energie van de proton bundels tijdens de tweede run van de Large Hadron Collider (LHC), waren de condities ideaal om het bestaan van supersymmetry aan de data te toetsen. Het eerste deel van deze thesis is toegewijd aan zoektochten naar sterke SUSY productie in evenementen met minimum 3 electrons of muons, jets, en ontbrekend momentum in het transversale vlak. Er zijn geen extra evenementen boven de SM achtergrond geobserveerd. Er zijn vier verschillende versimpelde SUSY modellen, die uitgaan van paar productie van gluinos of derde-generatie squarks, onderzocht. Voor elk van deze modellen zijn de parameterwaarden berekend waarvoor het model uitgesloten is met een 95% betrouwbaarheidsinterval. Voor het model met gluino paar productie, waarbij de gluinos vervolgens vervallen in een top quark-antiquark paar en een neutralino, kunnen we gluinos uitsluiten met een massa kleiner dan 1610 GeV wanneer we een veronderstellen dat het lichtste SUSY deeltje massaloos is. In het geval van het model voor bottom squark paar productie zijn de bottom squark massa's uitgesloten tot 840 GeV wanneer we veronderstellen dat de charginos lichter dan 200 GeV zijn. Voor het versimpelde model van top squark paar prodcutie, the $\widetilde{t}_2$ massa is uitgesloten voor waarden tot 720, 780, of 710 GeV in het geval van respectivelijk een exclusief $\widetilde{t}_2 \to \widetilde{t}_1 H$ verval, een exclusief $\widetilde{t}_2 \to \widetilde{t}_1 Z$ verval, of een mix waarin beide vervallen even vaak voorkomen.

In de nabije toekomst wordt er geen grote toenname meer verwacht in zowel de energie van de proton bundels als de te verzamelen data. Daarom moeten we op zoek gaan naar andere strategieën om nieuwe fysica buiten het SM te zoeken. Een mogelijkheid hiervoor is om een model-onafhankelijke methode toe te passen waarbij we zoeken naar nieuwe fysica in de veelvoorkomende electrozwakke processen, in de context van zware deeltjes zoals het BEH boson, de W en Z bosonen en de top quarks. Top quark paar productie in associatie met een massieve vector boson ($t\bar{t}V$) is hiervoor een uitstekende eerste keuze, gezien dit process relatief vaak voorkomt aan de LHC. Het tweede deel van deze thesis is dan ook gewijd aan de meting van de werkzame doorsnede van de $t\bar{t}W$ en $t\bar{t}Z$ processen. De meting van de $t\bar{t}W$ productie is uitgevoerd door een finale toestand van twee leptonen met dezelfde lading te selecteren in de 2016 data verzameld met de CMS detector, en hierbij werd een precisie van 20% bereikt, dankzij het gebruik van een multivariatie

techniek. Voor de meting van de $t\bar{t}Z$ productie werd zowel de CMS data uit 2016 en 2017 gebruikt. De werkzame doorsnede van het $t\bar{t}Z$ process is bepaald in evenementen met drie en vier leptonen, en bereikte een precisie van 8%. Het eindresultaat is het eerste experimenteel resultaat met een hogere precisie dan de state-of-the-art theoretische waarden die gebruik maken van next-to-leading-order berekeningen.

11

RÉSUMÉ

Le modèle standard de la physique des particules est l'une des théories les plus réussies du siècle dernier. Il résume notre compréhension de toutes les particules élémentaires connues et de nombreuses prédictions de cette théorie se sont confirmées au cours des dernières décennies. La découverte du boson BEH en 2012 par les collaborations CMS et ATLAS ont été le point culminant de la réussite de ce modèle. Alors qu'il décrit très bien les données expérimentales actuelles, l'indication de la teneur en matière noire dans l'univers, l'asymétrie matière-antimatière et les oscillations de neutrinos sont parmi les phénomènes qu'il ne peut prédire. En outre, il convient de mentionner que ce modèle ne couvre que trois des quatre forces fondamentales de la nature, à savoir les interactions électromagnétiques, faibles et fortes. La gravité, la force dominante à grande échelle, est tout simplement absente du modèle. Même s'il a réussi à prédire et à expliquer un large éventail de processus physiques, il ne peut pas être la théorie ultime décrivant l'univers.

Le grand volume de données et l'augmentation de l'énergie des faisceaux de protons du grand collisionneur de hadrons (LHC) lors de sa deuxième période d'exploitation ont pour but de démontrer l'existence ou non de la supersymétrie (SUSY). La première partie de cette thèse est consacrée à la recherche de particules supersymétriques dans les événements avec au moins 3 électrons ou muons, gerbes, et un manque d'énergie transverse. L'ensemble des données recueillies en 2016 ne fait état d'aucun dépassement significatif par rapport au bruit de fond prévu pour le modèle standard. Les limites d'exclusion avec un niveau de confiance de 95% sont calculées pour quatre modèles de supersymétrie simplifiés includant une production de paires de gluinos ou de squarks de troisième génération. Dans le modèle avec production de paires de gluinos, se décomposant ultérieurement en une paire quark-top quark anti-top et un neutrino, les gluinos ayant une masse inférieure à 1610 GeV sont exclus pour une particule supersymétrique la plus légère et n'ayant pas de masse. Dans le cas de la production de paires de squarks bottom, les masses de squark bottom inférieures à 840 GeV sont exclues pour les charginos de moins de 200 GeV. Pour un modèle simplifié de production de paires de squarks lourds, la masse $\widetilde{t}_2$ est exclue jusqu'à 720, 780, ou 710 GeV pour les modèles avec une désintégration exclusive $\widetilde{t}_2 \to \widetilde{t}_1 H$, $\widetilde{t}_2 \to \widetilde{t}_1 Z$, ou mélange équiprobable de ces deux désintégrations.

Dans un avenir proche, aucune augmentation significative de l'énergie des faisceaux de protons ou de la quantité de données collectées n'est attendue. Pour trouver la physique au-delà du modèle standard, nous devrons probablement réévaluer notre stratégie. Une possibilité est d'envisager une approche ne dépendant pas d'un modèle et de rechercher une nouvelle physique dans les processus électro-faibles, beaucoup plus abondants, qui impliquent des particules lourdes récemment découvertes comme les BEH, les bosons W et Z ou les quarks top. Le taux de production relativement élevé des paires de quarks top en association avec des bosons vecteurs massifs ($t\bar{t}V$) nous permet de les étudier en premier lieu. La deuxième partie de la thèse est consacrée à la mesure inclusive de la section transversale des processus $t\bar{t}W$ et $t\bar{t}Z$. La mesure de la production de $t\bar{t}W$

est effectuée dans l'état final de dilepton de même charge et a atteint un niveau de précision de 20% avec les données recueillies par l'expérience CMS en 2016. La pierre angulaire de cette mesure est l'utilisation d'une technique multivariée qui a permis d'atteindre le niveau de découverte du processus $t\bar{t}W$ dans les données à $\sqrt{s} = 13$ TeV. L'échantillon de données utilisé pour mesurer précisément la production de $t\bar{t}Z$ a été recueilli par l'expérience CMS en 2016 et 2017. La section efficace inclusive est mesurée dans les états finaux à trois et quatre leptons atteignant le niveau de précision de 8%. Le résultat rapporté est la première mesure expérimentale qui est plus précise que les calculs théoriques les plus récents estimés à NLO QCD et la précision électro-faible pour la production $t\bar{t}Z$.

A

LEPTON MVA MULTIVARIATE DISCRIMINATOR

The input variables of the lepton MVA used in the $t\bar{t}Z$ analysis and BDT score are inspected in di-lepton events in data. Three control regions are defined: one enriched in $Z \to \ell^+\ell^-$ events, another enriched in $t\bar{t} \to (be\nu)(\bar{b}\mu\nu)$ events, and a third region enriched in a semi-leptonic $t\bar{t}$ decay with an additional nonprompt lepton. Events are selected with a mix of isolated single and double lepton triggers and to ensure high efficiency of the triggering the highest-p_T and second highest-p_T leptons have to pass 25 and 15 GeV requirement in selection. The leptons are selected using quite loose requirement on impact parameters and isolation, and muons were additionally required to pass the "Medium ID" requirement. The presence of a third lepton is vetoed. In each plot, the total yield from simulation is normalized to the events number in data, to facilitate the comparison of the simulated shapes to the data in each distribution.

All plots in this chapter are courtesy of Willem Verbeke.

a.1 CONTROL REGION ENRICHED IN DY EVENTS

The DY control region is obtained by selecting events with two leptons of opposite charge and same flavor ($\mu^+\mu^-$, e^+e^-) that have a mass 10 GeV close to the Z boson mass; additionally we require no b jets to be present in the event.

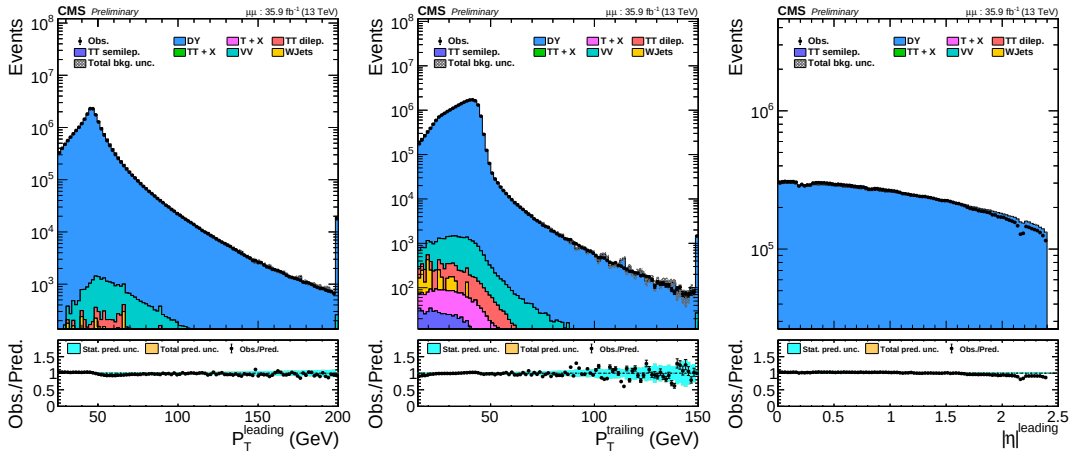


Figure A.1: p_T of the leading lepton (right), p_T of the trailing lepton (middle), and $|\eta|$ of the leading lepton in 2016 $\mu\mu$ data.

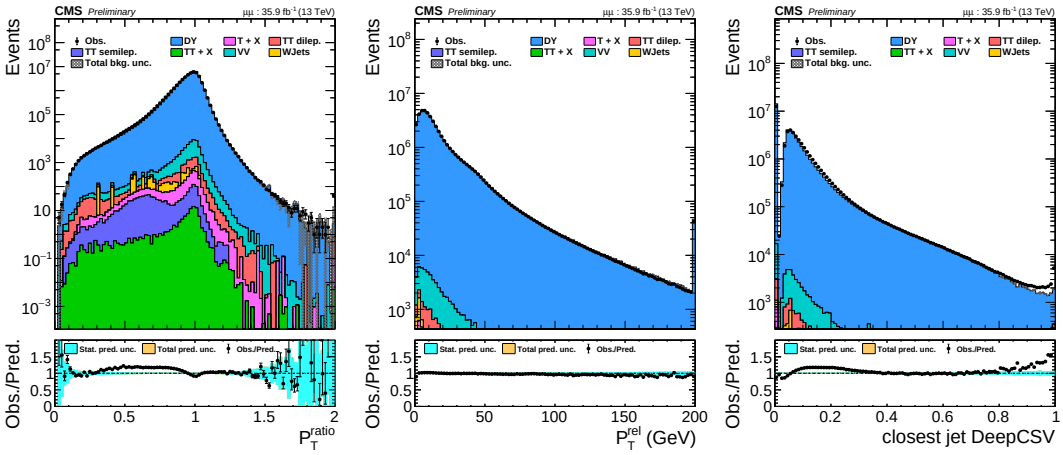


Figure A.2: $p_{\mathrm{T}}^{\mathrm{ratio}}$ (right), $p_{\mathrm{T}}^{\mathrm{rel}}$ and DeepCSV b tagger value in 2016 $\mu\mu$ data.

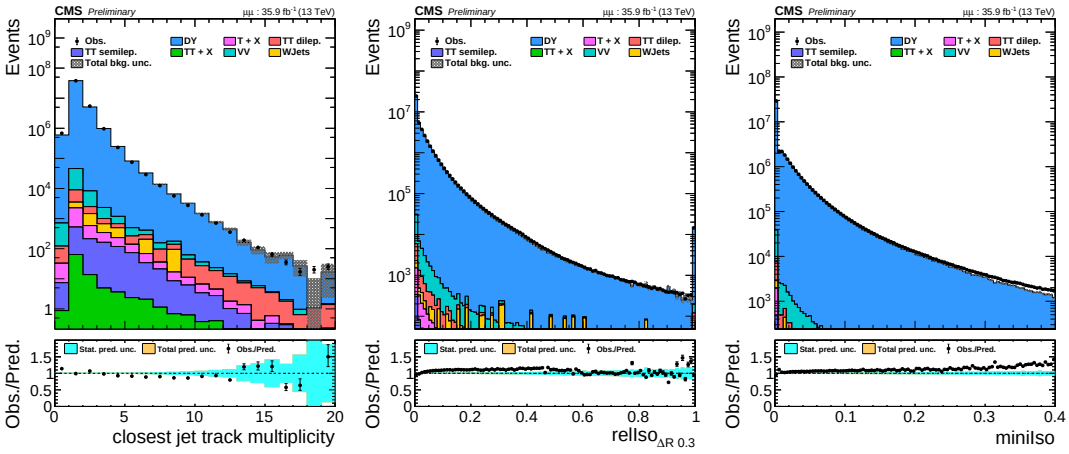


Figure A.3: Track multiplicity of the closest jet, I_{rel}, and I_{mini} in 2016 $\mu\mu$ data.

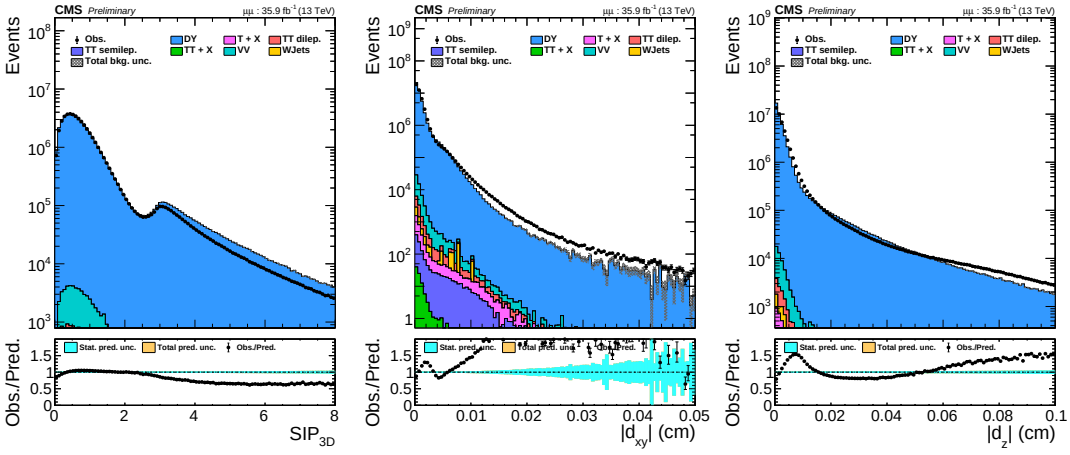


Figure A.4: $\mathrm{SIP_{3D}}$, $\mathrm{d_{xy}}$ and $\mathrm{d_z}$ in 2016 $\mu\mu$ data.

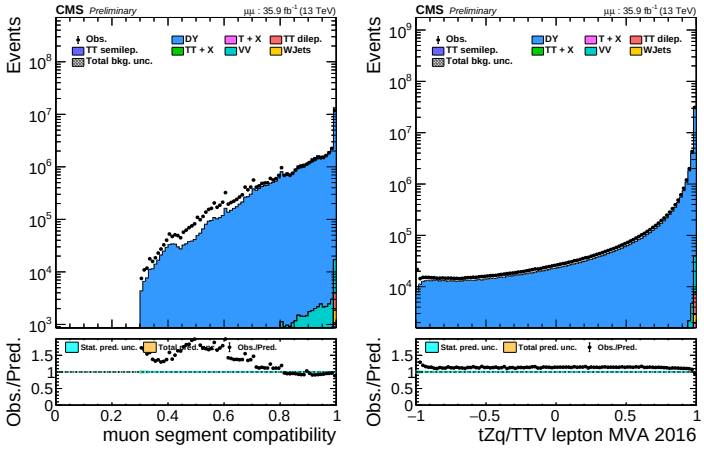


Figure A.5: Muon segment compatibility and Lepton MVA output in 2016 $\mu\mu$ data.

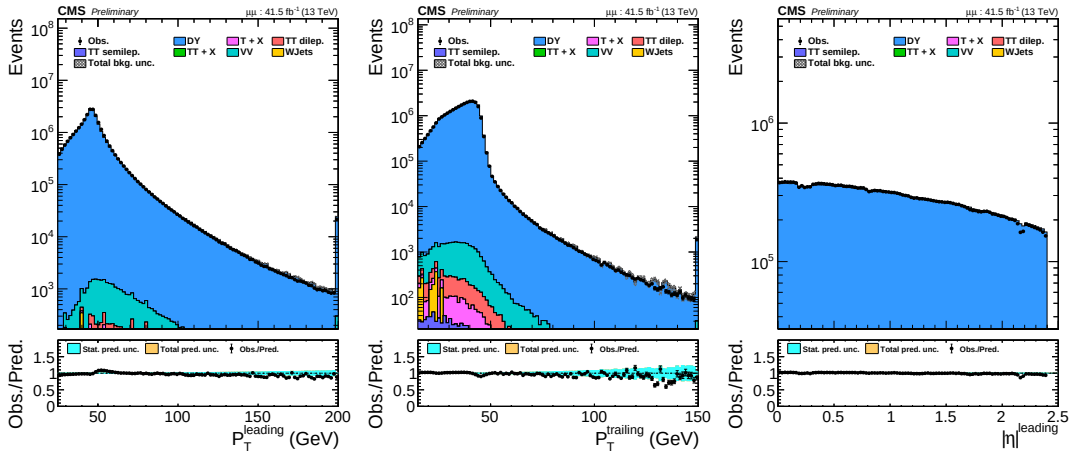


Figure A.6: p_{T} of the leading lepton (right), p_{T} of the trailing lepton (middle), and $|\eta|$ of the leading lepton in 2017 $\mu\mu$ data.

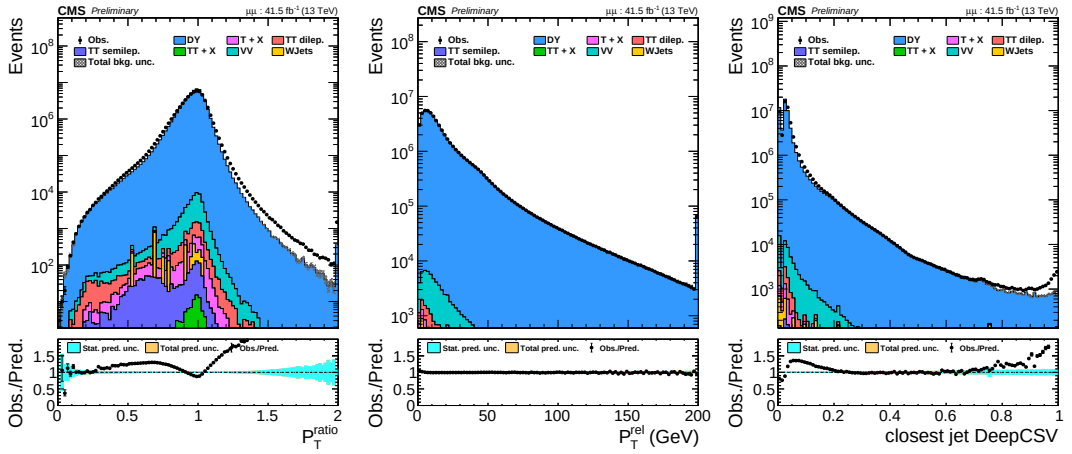


Figure A.7: $p_{\mathrm{T}}^{\mathrm{ratio}}$ (right), $p_{\mathrm{T}}^{\mathrm{rel}}$ and DeepCSV b tagger value in 2017 $\mu\mu$ data.

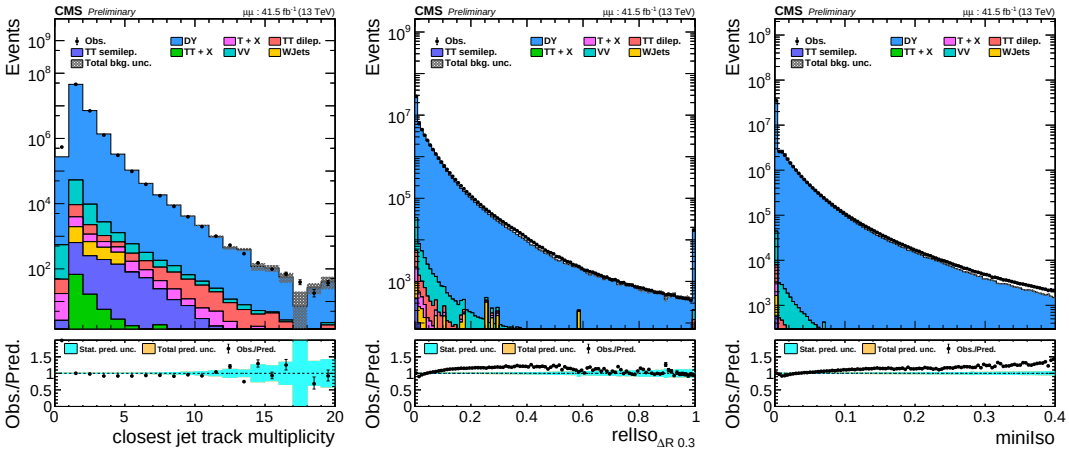


Figure A.8: Track multiplicity of the closest jet, I_{rel}, and I_{mini} in 2017 $\mu\mu$ data.

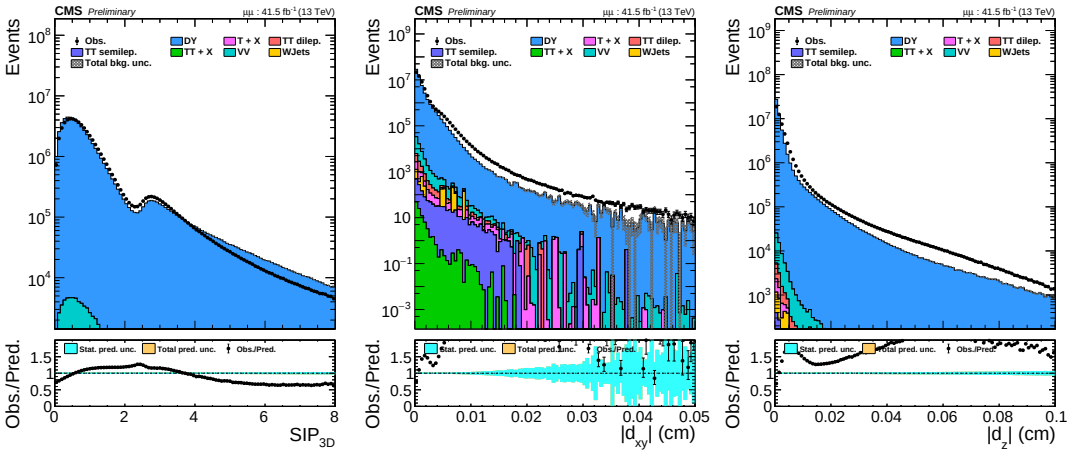


Figure A.9: SIP_{3D}, d_{xy} and d_{z} in 2017 $\mu\mu$ data.

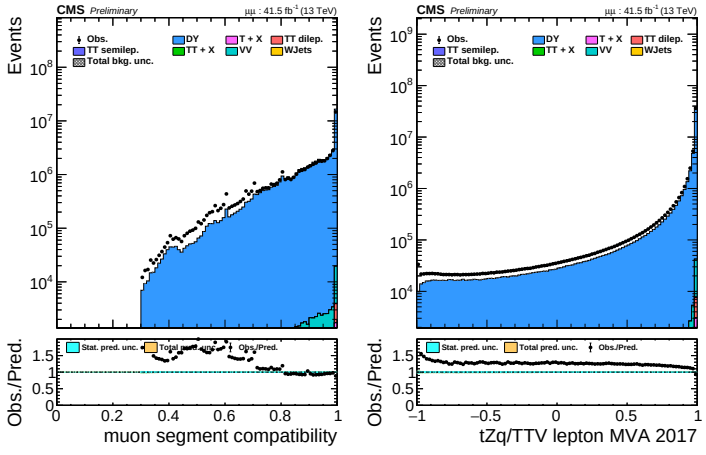


Figure A.10: Muon segment compatibility and Lepton MVA output in 2017 $\mu\mu$ data.

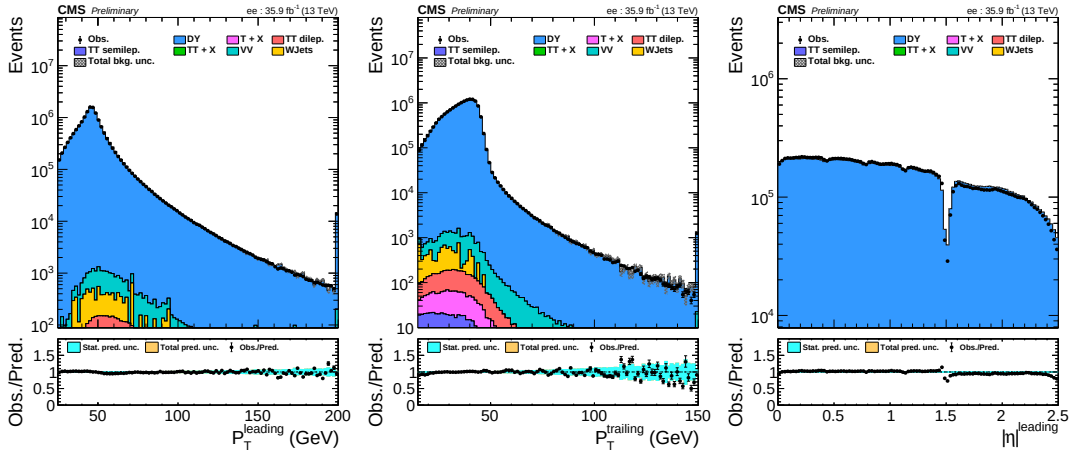


Figure A.11: p_T of the leading lepton (right), p_T of the trailing lepton (middle), and $|\eta|$ of the leading lepton in 2016 *ee* data.

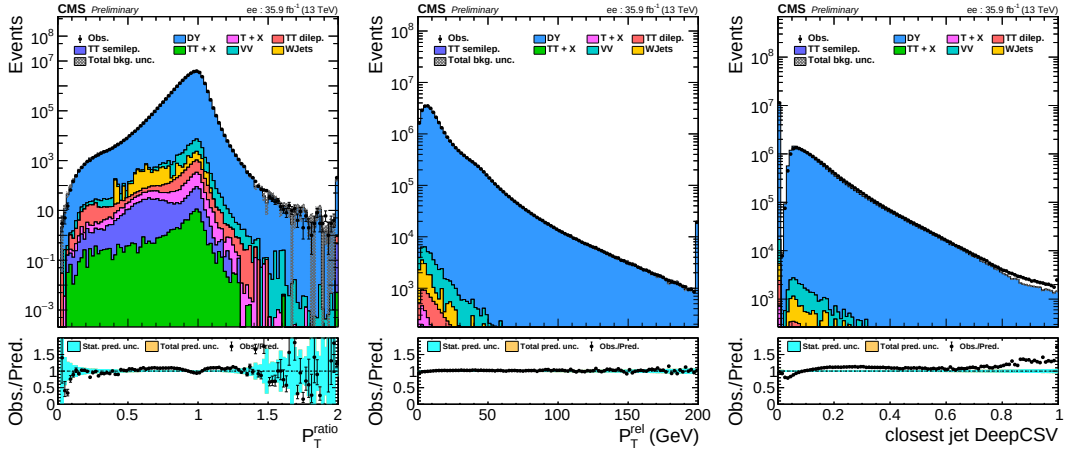


Figure A.12: p_T^{ratio} (right), p_T^{rel} and DeepCSV b tagger value in 2016 *ee* data.

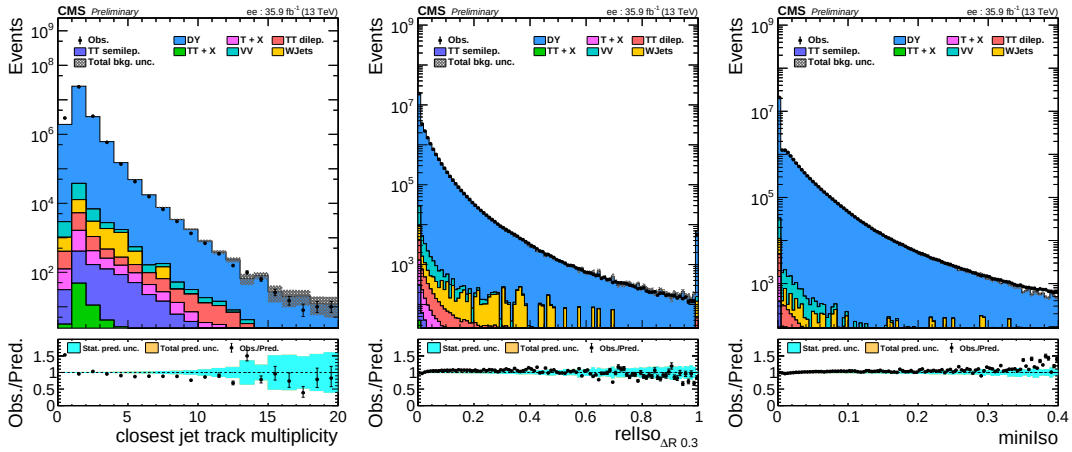


Figure A.13: Track multiplicity of the closest jet, I_{rel}, and I_{mini} in 2016 *ee* data.

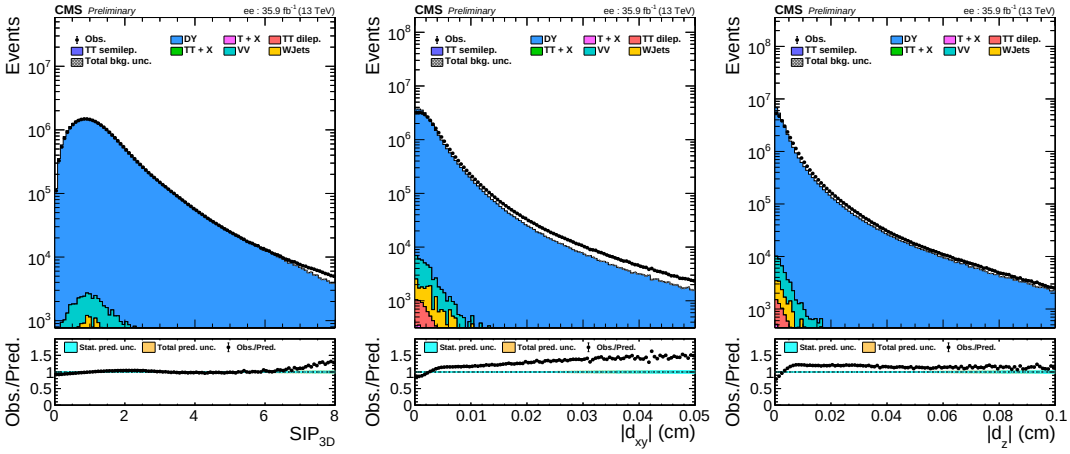


Figure A.14: $\mathrm{SIP_{3D}}$, $\mathrm{d_{xy}}$ and $\mathrm{d_z}$ in 2016 *ee* data.

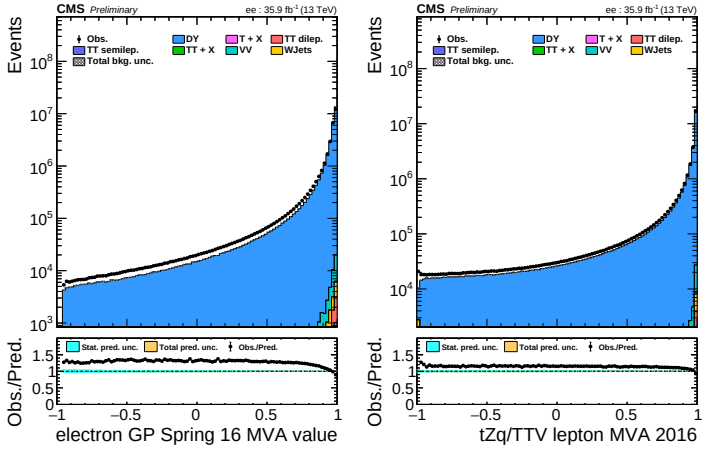


Figure A.15: Electron MVA and Lepton MVA output in 2016 *ee* data.

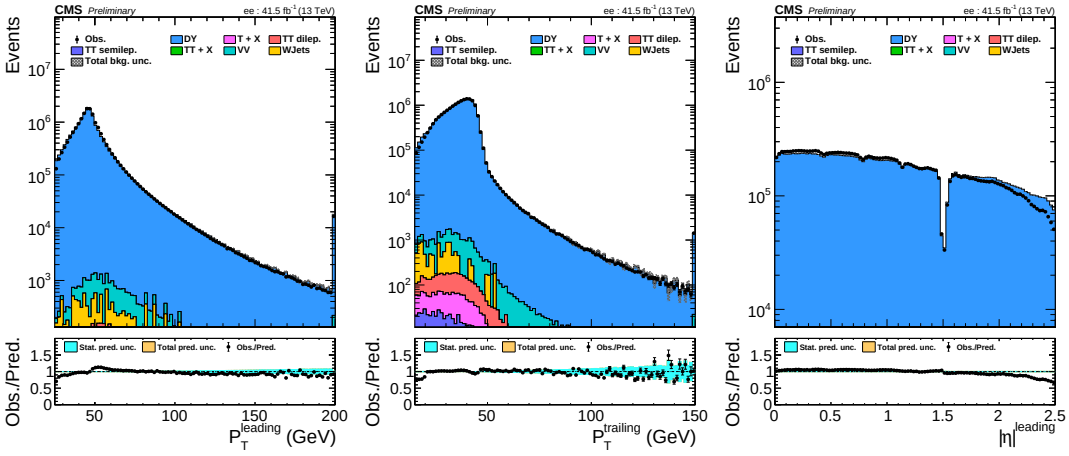


Figure A.16: p_T of the leading lepton (right), p_T of the trailing lepton (middle), and $|\eta|$ of the leading lepton in 2017 *ee* data.

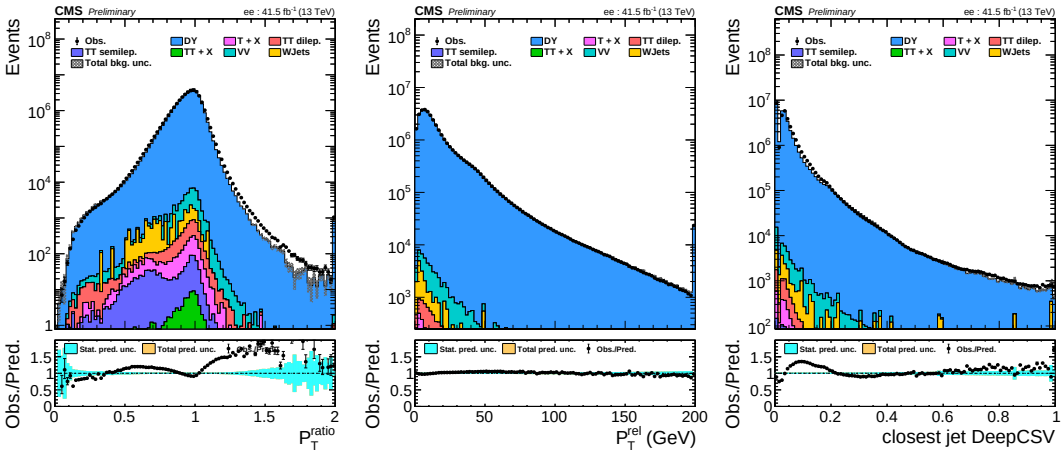


Figure A.17: $p_{\mathrm{T}}^{\mathrm{ratio}}$ (right), $p_{\mathrm{T}}^{\mathrm{rel}}$ and DeepCSV b tagger value in 2017 *ee* data.

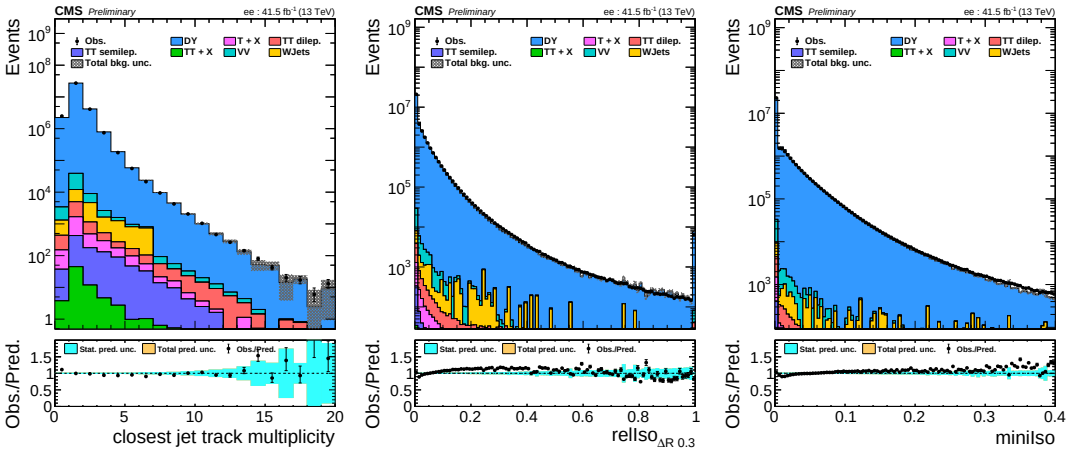


Figure A.18: Track multiplicity of the closest jet, I_{rel}, and I_{mini} in 2017 *ee* data.

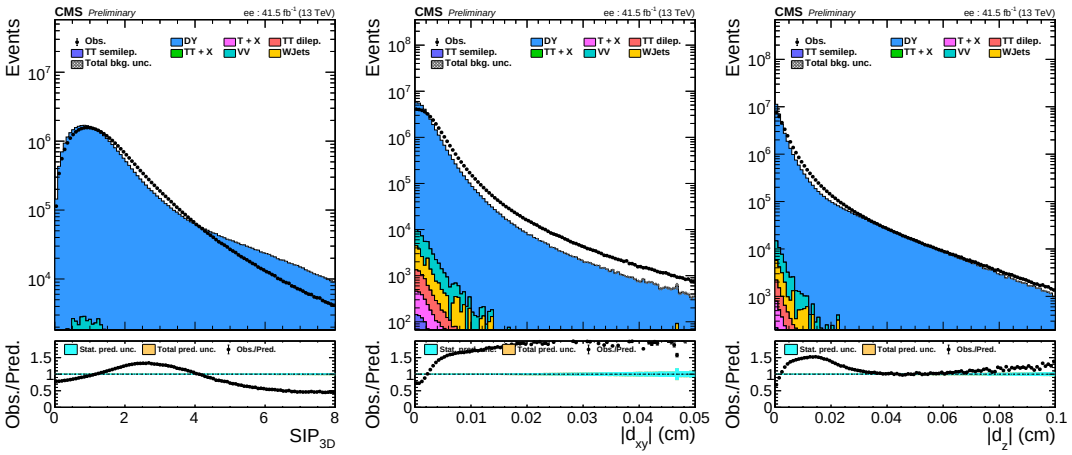


Figure A.19: $\mathrm{SIP_{3D}}$, $\mathrm{d_{xy}}$ and $\mathrm{d_z}$ in 2017 *ee* data.

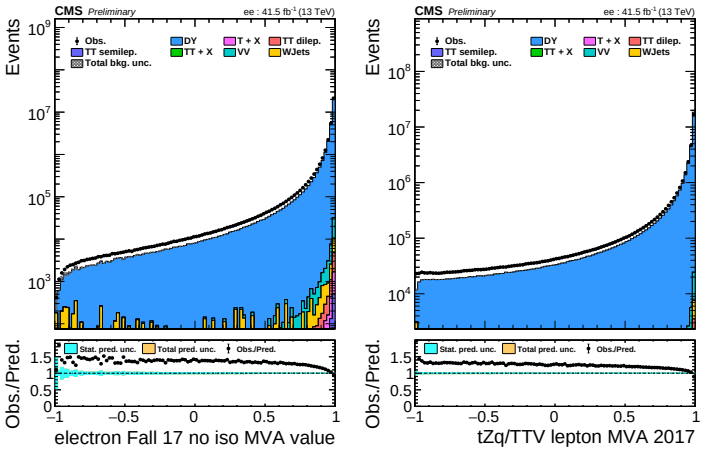


Figure A.20: Electron MVA and Lepton MVA output in 2017 *ee* data.

a.2 CONTROL REGION ENRICHED IN TWO PROMPT LEPTON EVENTS

Additionally to the requirement on opposite flavor pair in di-leptonic $t\bar{t}$ enriched region, the presence of at least 2 jets with $|\eta| < 2.4$ and at least 1 b jet is required. In addition, p_{T}^{miss} is required to be greater than 50 GeV in the $t\bar{t}$ enriched region.

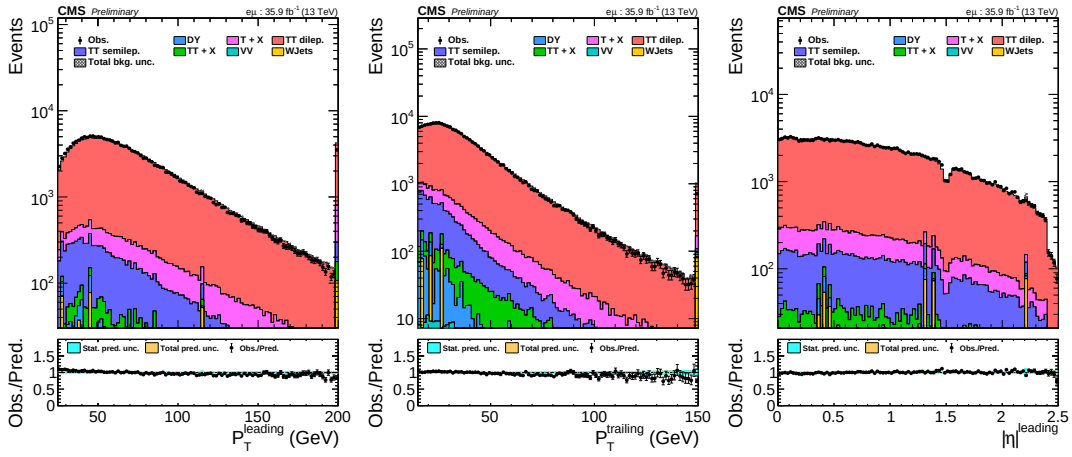


Figure A.21: p_{T} of the leading lepton (right), p_{T} of the trailing lepton (middle), and $|\eta|$ of the leading lepton in 2016 $e\mu$ data.

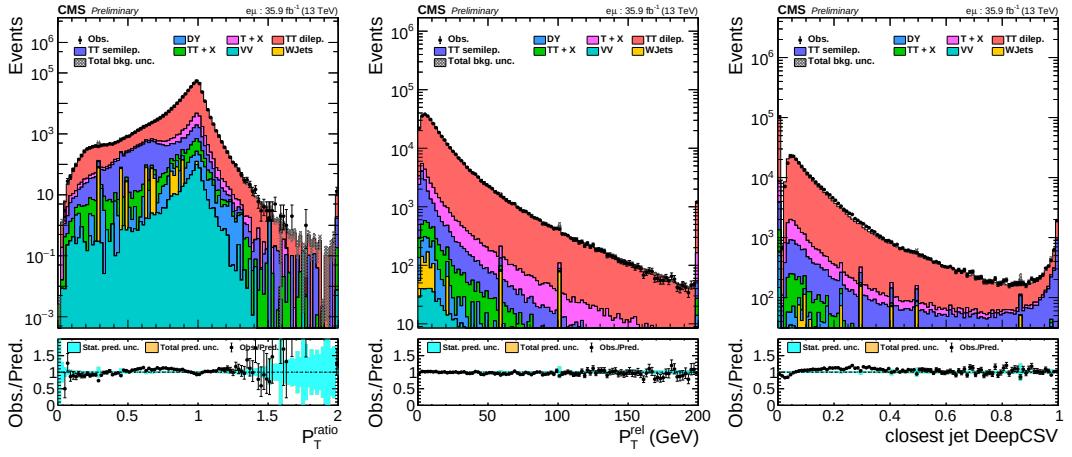


Figure A.22: $p_{\mathrm{T}}^{\mathrm{ratio}}$ (right), $p_{\mathrm{T}}^{\mathrm{rel}}$ and DeepCSV b tagger value in 2016 $e\mu$ data.

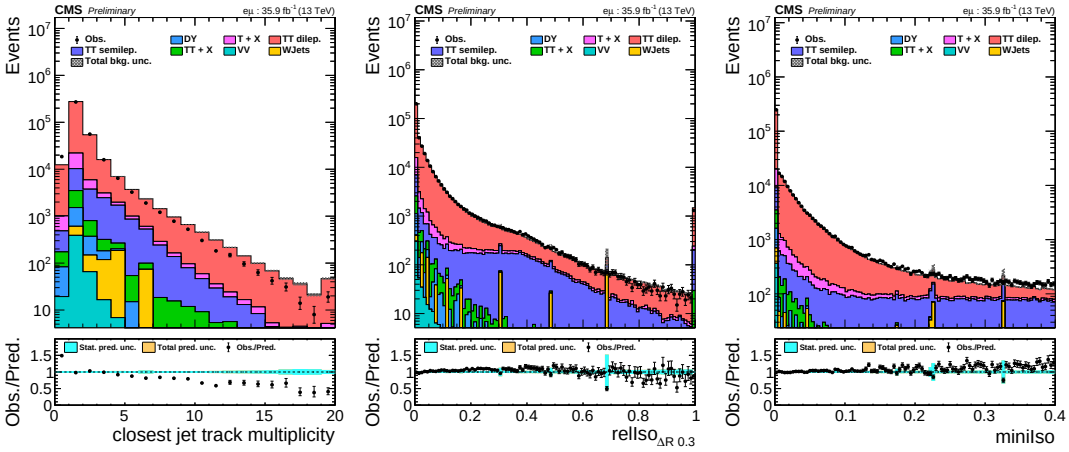


Figure A.23: Track multiplicity of the closest jet, I_{rel}, and I_{mini} in 2016 $e\mu$ data.

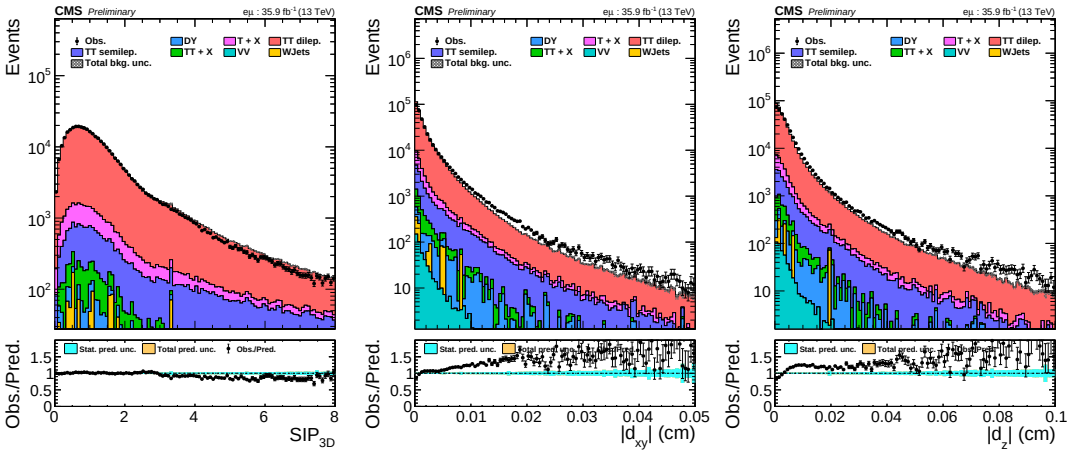


Figure A.24: SIP_{3D}, d_{xy} and d_{z} in 2016 $e\mu$ data.

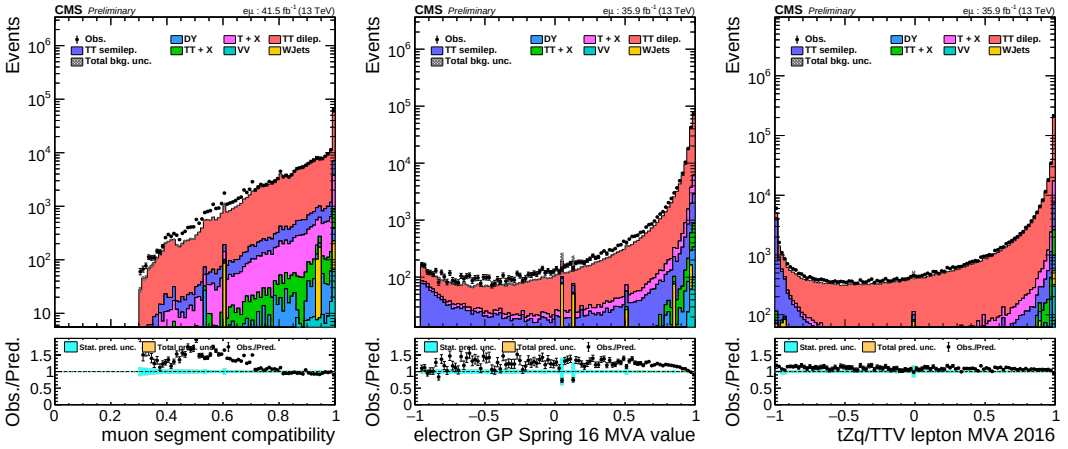


Figure A.25: Muon segment compatibility (left), electron MVA (middle) and Lepton MVA output (right) in 2016 $e\mu$ data.

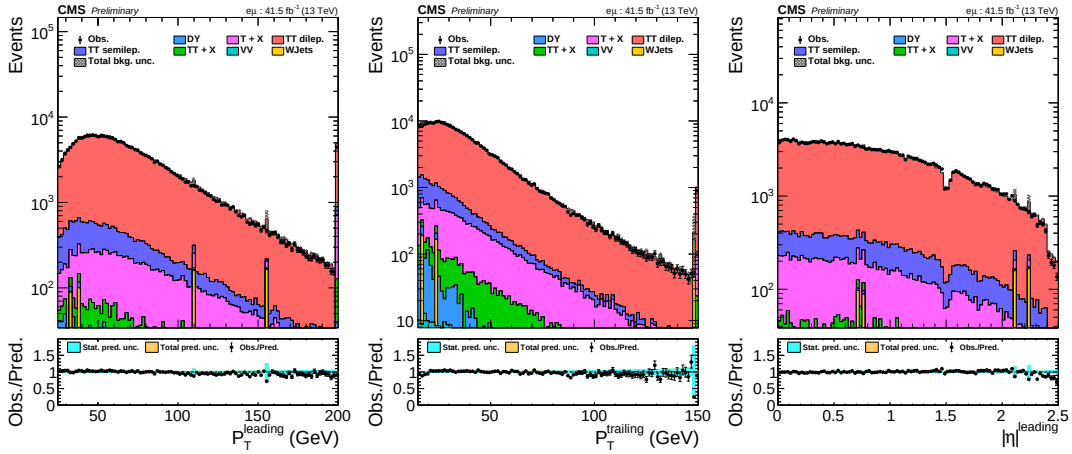


Figure A.26: p_{T} of the leading lepton (right), p_{T} of the trailing lepton (middle), and $|\eta|$ of the leading lepton in 2017 $e\mu$ data.

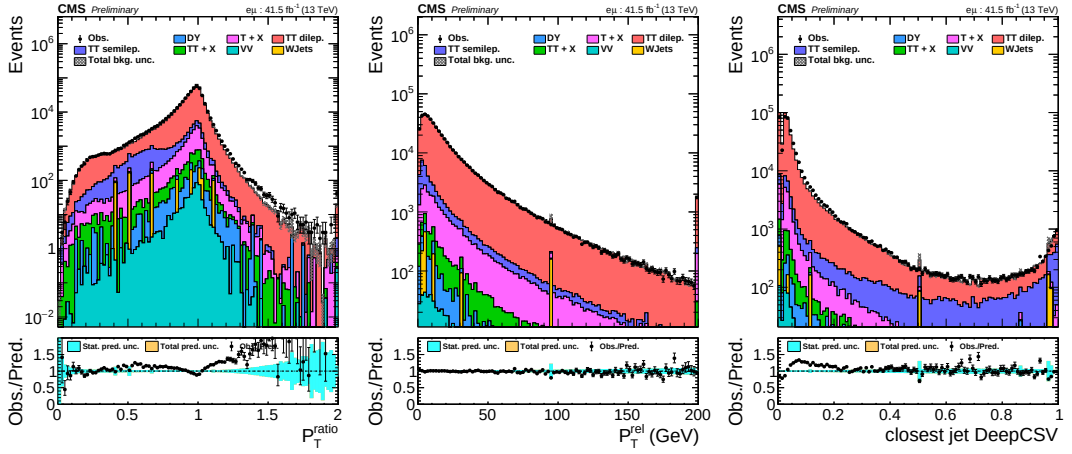


Figure A.27: $p_{\mathrm{T}}^{\mathrm{ratio}}$ (right), $p_{\mathrm{T}}^{\mathrm{rel}}$ and DeepCSV b tagger value in 2017 $e\mu$ data.

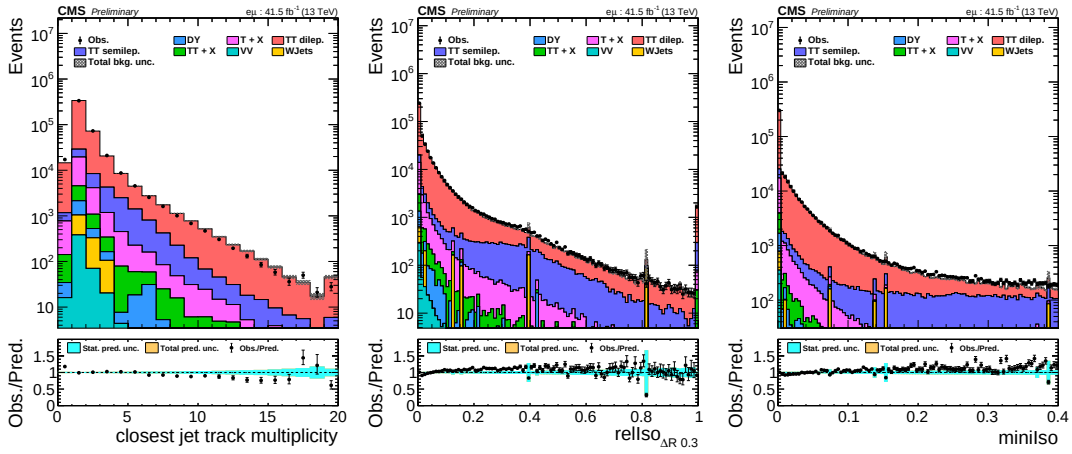


Figure A.28: Track multiplicity of the closest jet, I_{rel}, and I_{mini} in 2017 $e\mu$ data.

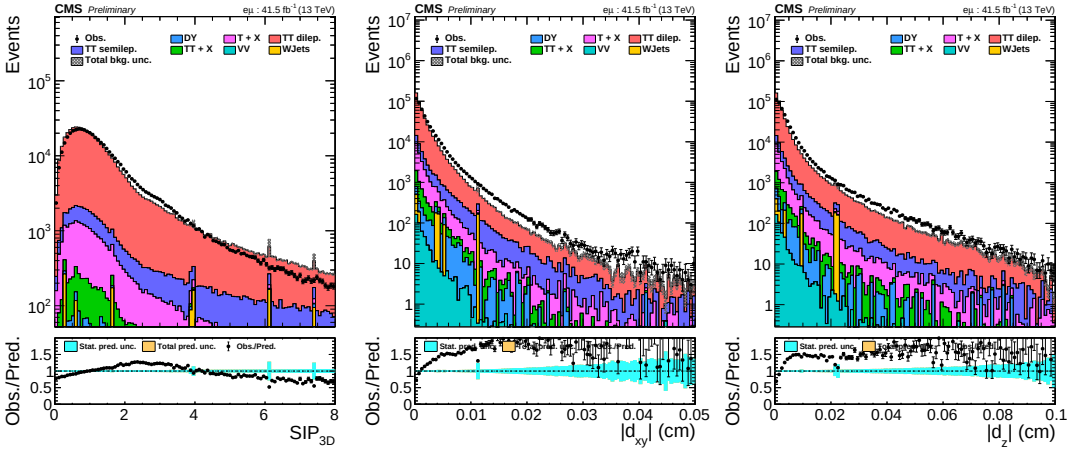


Figure A.29: SIP_{3D}, d_{xy} and d_z in 2017 $e\mu$ data.

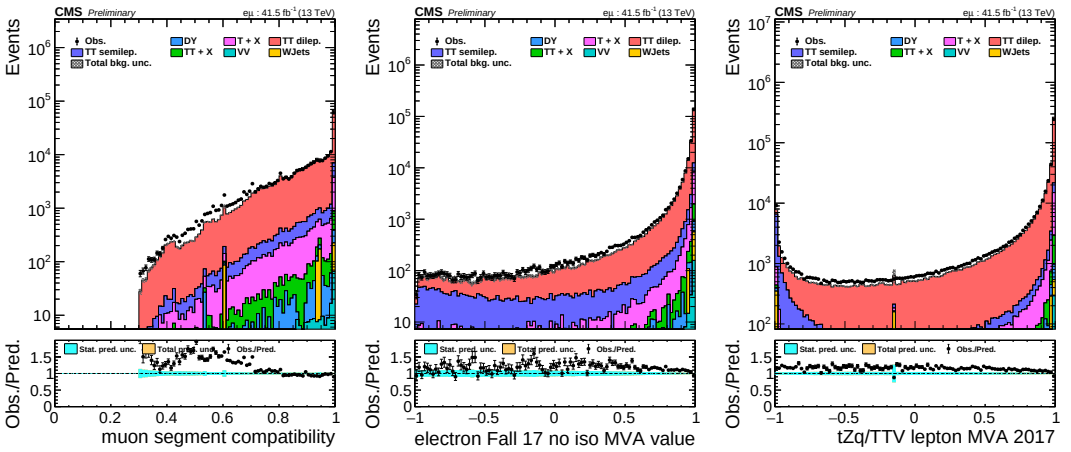


Figure A.30: Muon segment compatibility (left), electron MVA (middle) and Lepton MVA output (right) in 2017 $e\mu$ data.

a.3 CONTROL REGION ENRICHED IN ONE PROMPT AND ONE NONMPROMPT LEPTON EVENTS

Finally, a control sample enriched in nonprompt leptons is constructed by requiring 2 same charge leptons, and exactly 3 or 4 jets with $|\eta| < 2.4$ and exactly 1 b jet.

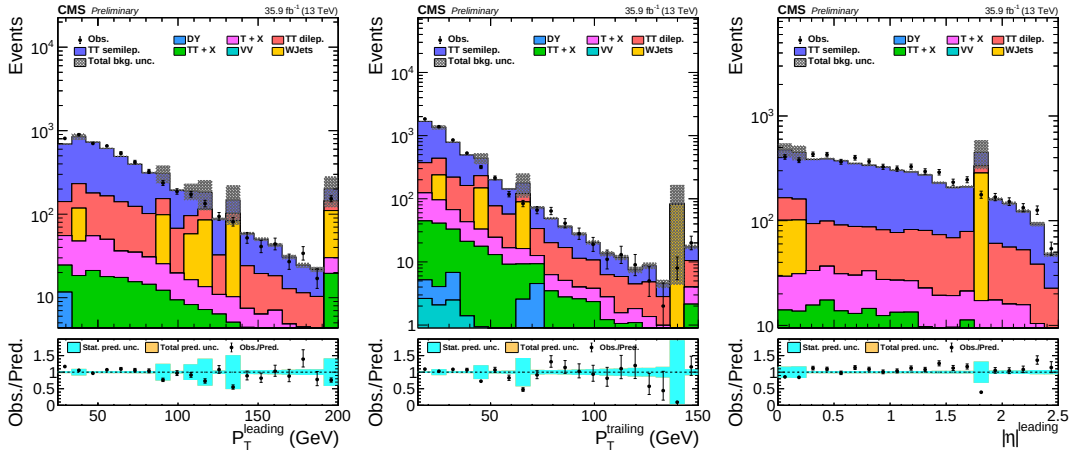


Figure A.31: p_{T} of the leading lepton (right), p_{T} of the trailing lepton (middle), and $|\eta|$ of the leading lepton in 2016 same-charge $e\mu$ data.

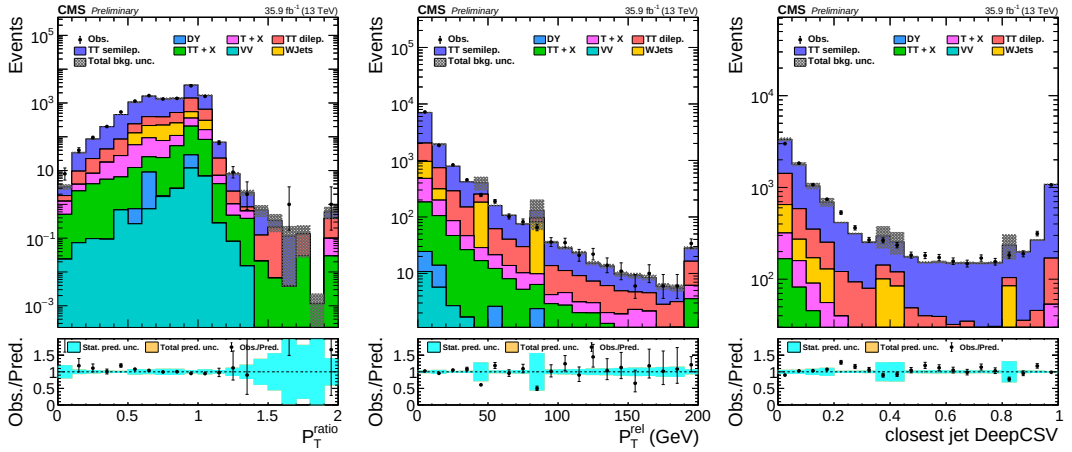


Figure A.32: $p_{\mathrm{T}}^{\mathrm{ratio}}$ (right), $p_{\mathrm{T}}^{\mathrm{rel}}$ and DeepCSV b tagger value in 2016 same-charge $e\mu$ data.

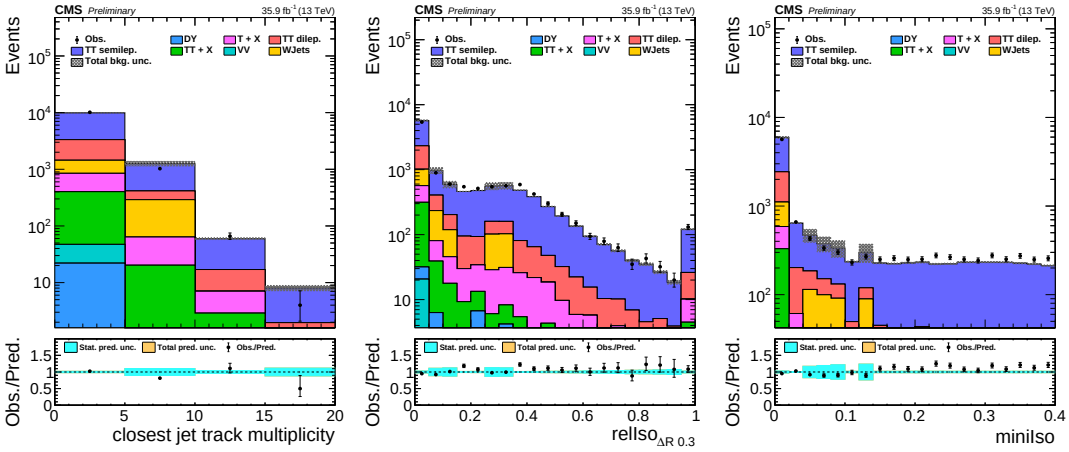


Figure A.33: Track multiplicity of the closest jet, I_{rel}, and I_{mini} in 2016 same-charge $e\mu$ data.

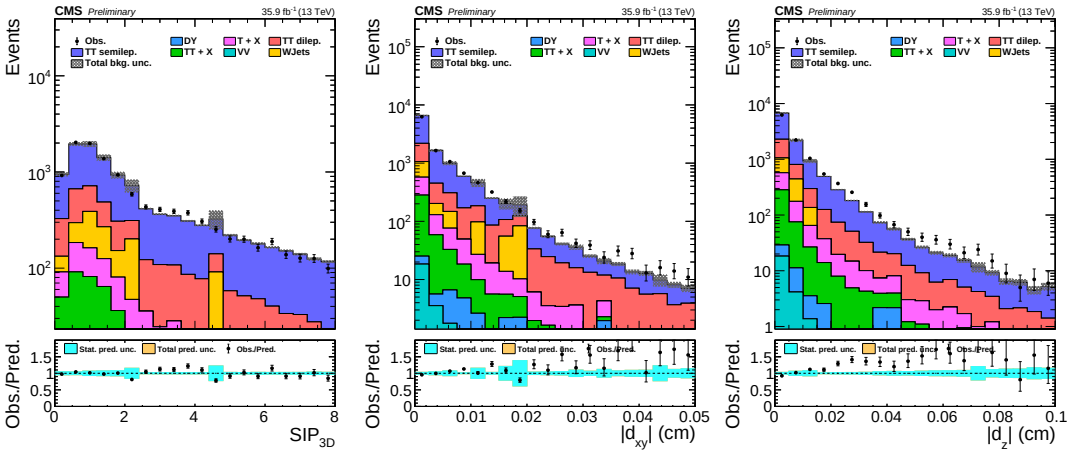


Figure A.34: $\text{SIP}_{3\text{D}}$, d_{xy} and d_{z} in 2016 same-charge $e\mu$ data.

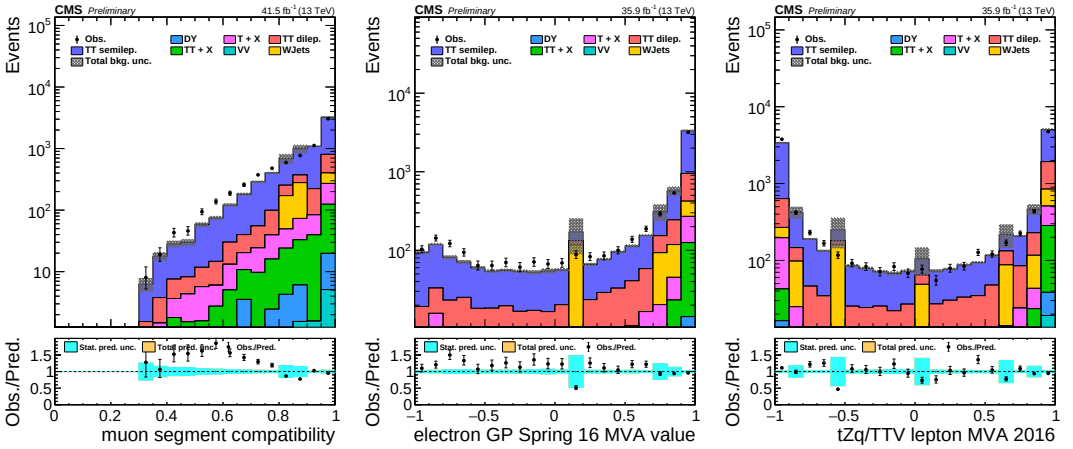


Figure A.35: Muon segment compatibility (left), electron MVA (right) and Lepton MVA output (right) in 2016 same-charge $e\mu$ data.

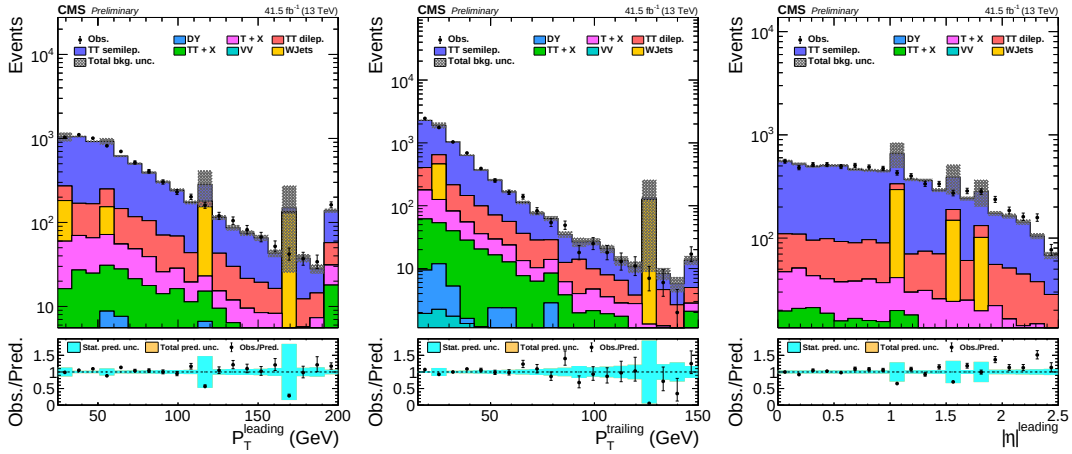


Figure A.36: p_{T} of the leading lepton (right), p_{T} of the trailing lepton (middle), and $|\eta|$ of the leading lepton in 2017 same-charge $e\mu$ data.

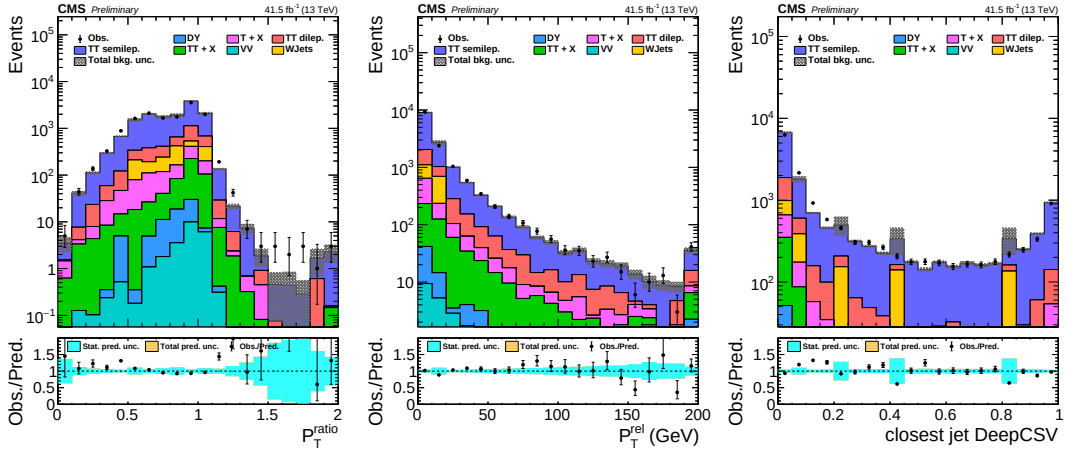


Figure A.37: $p_{\mathrm{T}}^{\mathrm{ratio}}$ (right), $p_{\mathrm{T}}^{\mathrm{rel}}$ and DeepCSV b tagger value in 2017 same-charge $e\mu$ data.

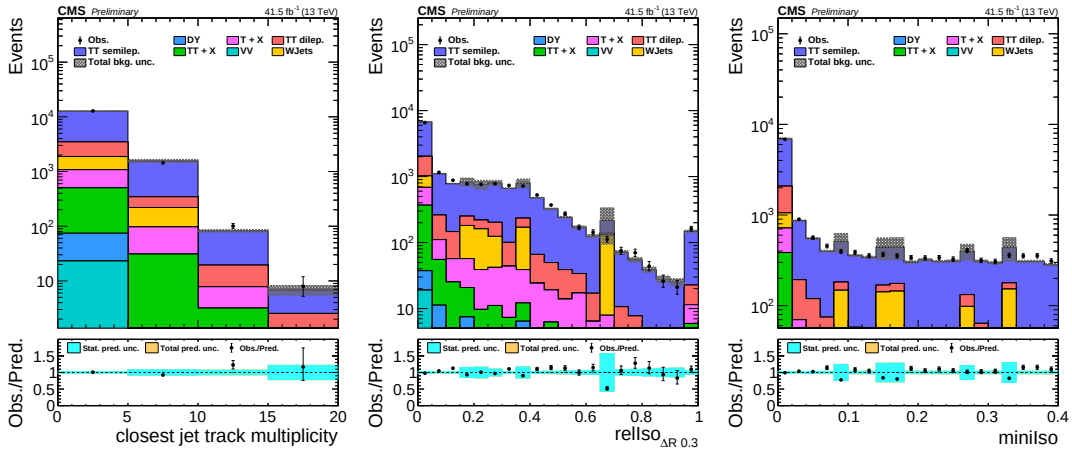


Figure A.38: Track multiplicity of the closest jet, I_{rel}, and I_{mini} in 2017 same-charge $e\mu$ data.

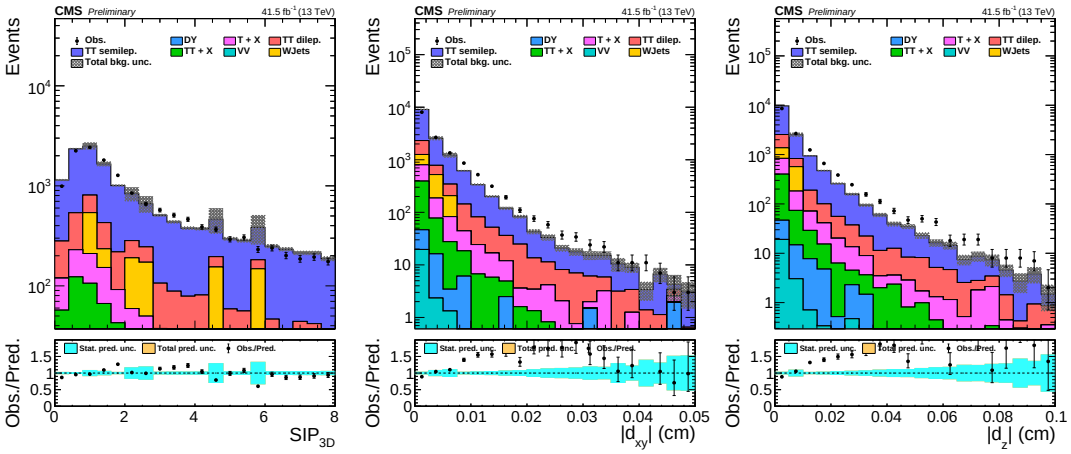


Figure A.39: SIP_{3D}, d_{xy} and d_z in 2017 same-charge $e\mu$ data.

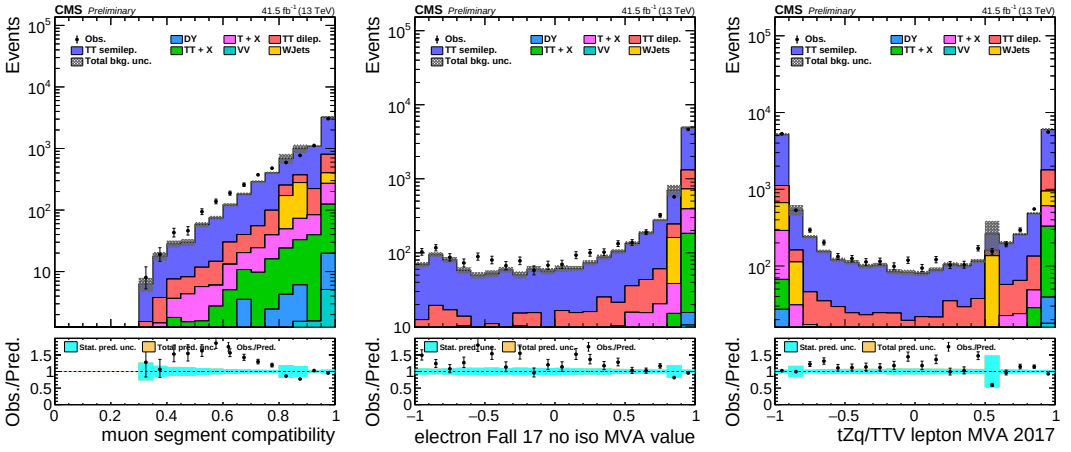


Figure A.40: Muon segment compatibility (left), electron MVA (middle) and Lepton MVA output (right) in 2017 same-charge $e\mu$ data.

B

SEARCH FOR SUPERSYMMETRY IN EVENTS WITH MULTIPLE CHARGED LEPTONS, JETS, AND MISSING TRANSVERSE MOMENTUM

b.1 HIGH-LEVEL TRIGGER PATHS

Table B.1 lists the high-level trigger paths that are used to select signal events in SUSY analysis. All of these dilepton triggers have not been prescaled during the relevant data taking period. Auxiliary single lepton triggers, which are listed in Table B.2, are used to select QCD multijet events for measuring the lepton misidentification rate, which is used for the estimation of the nonprompt lepton background. The auxiliary triggers have been prescaled to reduce their rate.

	Channel	Trigger Name
isolated	$\mu\mu$	HLT_Mu17_TrkIsoVVL_(Tk)Mu8_TrkIsoVVL(_DZ)
	ee	HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL_DZ
	$e\mu$	HLT_Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL(_DZ) HLT_Mu23_TrkIsoVVL_Ele8_CaloIdL_TrackIdL_IsoVL(_DZ)
non-isolated	$\mu\mu$	HLT_DoubleMu8_Mass8_PFHT300
	ee	HLT_DoubleEle8_CaloIdM_TrackIdM_Mass8_PFHT300
	$e\mu$	HLT_Mu8_Ele8_CaloIdM_TrackIdM_Mass8_PFHT300

Table B.1: Set of triggers used in the SUSY analysis. The requirement on the distance between two leptons (indicated in blue in the names) is used only in the dataset collected in the latest run era and corresponds to approximately 9 fb^{-1} of the total collected data in 2016. For this period the triggers without mentioned requirement were prescaled.

Table B.2: Auxiliary triggers used to select QCD multijet events for measuring the lepton misidentification rate.

Channel	Trigger Name
μ	HLT_Mu8 HLT_Mu17
e	HLT_Ele8_CaloIdM_TrackIdM_PFJet30 HLT_Ele17_CaloIdM_TrackIdM_PFJet30

b.2 LIST OF SM SIGNAL AND BACKGROUND SIMULATIONS

Table B.3 lists all MC simulations for SM signal and background processes used in the SUSY analysis. The cross sections used for the normalizations of the background MC samples are calculated from the matrix element of the MC generator used. The cross sections for signal samples are calculated at NLO-NLL(next-to-leading-logarithm) accuracy and assumes that other SUSY particles are decoupled (i.e. very massive). The NNPDF3.0 [120, 137] PDFs are used for the hard scattering process in the samples. The modeling of the underlying event is generated using the CUETP8M1 [131, 138] for most of the samples in signal and background categories, apart from the $t\bar{t}H$ sample where the CUETP8M2 tune [139] is utilized.

The background simulations listed in the 'nonprompt' category are not used for the final result of the analysis where this type of background is estimated with a data-driven technique. Instead, these simulations are used for closure tests to validate the nonprompt lepton background estimation.

Category	Sample Name	Generator	Matching scheme	σ pb
Signal	-	MADGRAPH5_aMC@NLO	MLM	-
nonprompt	$t\bar{t}$ to 2ℓ	Madgraph	MLM	87.315
	$t\bar{t}$ to 1ℓ	Madgraph	MLM	364.36
	DY to 2ℓ ($m_{\ell\ell} > 50$ GeV)	MADGRAPH5_aMC@NLO	FxFx	6024
	DY to 2ℓ (10 GeV $< m_{\ell\ell} < 50$ GeV)	MADGRAPH5_aMC@NLO	FxFx	18610
WZ	WZ to $3\ell 1\nu$	Powheg	-	4.4297
$t\bar{t}Z$	$t\bar{t}Z$, Z to 2ℓ or 2ν ($m_{\ell\ell} > 10$ GeV)	MADGRAPH5_aMC@NLO	-	0.2529
	$t\bar{t}Z$, Z to 2ℓ (1 GeV $< m_{\ell\ell} < 10$ GeV)	Madgraph	MLM	0.0283
$t\bar{t}X$	$t\bar{t}W$, W to $1\ell 1\nu$	MADGRAPH5_aMC@NLO	FxFx	0.2043
	$t\bar{t}H$, H to all, except $b\bar{b}$	Powheg	-	0.2151
	tZq, Z to 2ℓ	MADGRAPH5_aMC@NLO	-	0.0758
	tHq, H inclusive	Madgraph	-	0.261
	tHW, H inclusive	Madgraph	-	0.01561
	tWZ, Z to 2ℓ	Madgraph	-	0.01123
	$t\bar{t}t\bar{t}$	MADGRAPH5_aMC@NLO	-	0.009103
	$t\bar{t}\gamma$	MADGRAPH5_aMC@NLO	FxFx	3.697
	$t\gamma$	MADGRAPH5_aMC@NLO	-	2.967
Rare SM	ZZ, both Z to 2ℓ	Powheg	-	1.256
	HZZ, both Z to 2ℓ	Powheg	-	0.0121
	VH, H to all, except $b\bar{b}$	MADGRAPH5_aMC@NLO	FxFx	0.9561
	ZZZ	MADGRAPH5_aMC@NLO	-	0.01398
	WZZ	MADGRAPH5_aMC@NLO	-	0.05565
	WWZ	MADGRAPH5_aMC@NLO	-	0.1651
	WWW	MADGRAPH5_aMC@NLO	-	0.2086
	$Z\gamma$	MADGRAPH5_aMC@NLO	FxFx	131.3
	$W\gamma$	MADGRAPH5_aMC@NLO	MLM	585.5

Table B.3: List of MC simulations of standard model background processes used in the SUSY analysis. Category name, sample name, generator name, the parton matching scheme and processes cross section are presented.

b.3 YIELDS FOR MC SIMULATION IN EACH SR

Table B.4: Simulated yields for 35.9 fb^{-1} of the rare SM backgrounds in the 23 on-Z signal regions and the associated statistical uncertainty from the size of the samples. Shown uncertainty is MC statistics.

	$t\bar{t}Z$	$t\bar{t}X$	WZ	Rares	total
SR1a	18.47 ± 0.69	12.22 ± 0.27	189.80 ± 3.83	20.88 ± 1.35	241.37 ± 4.12
SR1b	3.11 ± 0.29	1.82 ± 0.16	13.17 ± 1.00	2.65 ± 0.26	20.75 ± 1.09
SR2a	4.73 ± 0.35	2.41 ± 0.12	38.72 ± 1.73	4.03 ± 0.35	49.89 ± 1.80
SR2b	1.05 ± 0.18	0.51 ± 0.08	2.03 ± 0.40	0.57 ± 0.12	4.17 ± 0.46
SR3a	4.48 ± 0.39	1.41 ± 0.10	30.72 ± 1.55	4.29 ± 0.60	40.90 ± 1.71
SR3b	0.54 ± 0.14	0.17 ± 0.04	3.31 ± 0.51	0.59 ± 0.09	4.63 ± 0.53
SR4a	1.59 ± 0.21	0.75 ± 0.08	12.12 ± 0.97	0.75 ± 0.13	15.22 ± 1.01
SR4b	0.16 ± 0.09	0.12 ± 0.03	0.87 ± 0.26	0.14 ± 0.10	1.29 ± 0.30
SR5	41.56 ± 0.95	20.87 ± 0.40	16.16 ± 1.12	3.14 ± 0.78	81.73 ± 1.71
SR6	9.93 ± 0.47	4.18 ± 0.18	3.57 ± 0.53	0.63 ± 0.13	18.32 ± 0.74
SR7	13.21 ± 0.60	3.59 ± 0.16	3.91 ± 0.56	0.68 ± 0.15	21.40 ± 0.85
SR8	3.59 ± 0.31	1.19 ± 0.10	1.48 ± 0.35	0.14 ± 0.04	6.40 ± 0.48
SR9	27.79 ± 0.75	9.81 ± 0.27	1.63 ± 0.34	0.55 ± 0.09	39.78 ± 0.87
SR10	4.34 ± 0.30	1.28 ± 0.10	0.14 ± 0.10	0.10 ± 0.04	5.86 ± 0.33
SR11	8.00 ± 0.45	1.69 ± 0.11	0.17 ± 0.12	0.16 ± 0.07	10.01 ± 0.48
SR12	1.58 ± 0.21	0.47 ± 0.05	0.15 ± 0.11	0.01 ± 0.01	2.21 ± 0.24
SR13	3.37 ± 0.27	0.70 ± 0.06	0.06 ± 0.06	0.02 ± 0.01	4.15 ± 0.28
SR14a	10.30 ± 0.65	2.08 ± 0.12	11.29 ± 0.94	5.07 ± 1.09	28.74 ± 1.58
SR14b	1.53 ± 0.26	0.53 ± 0.06	1.76 ± 0.37	0.54 ± 0.12	4.36 ± 0.47
SR15a	4.41 ± 0.40	1.33 ± 0.10	7.62 ± 0.77	1.10 ± 0.15	14.46 ± 0.89
SR15b	0.69 ± 0.18	0.45 ± 0.06	0.49 ± 0.20	0.15 ± 0.05	1.78 ± 0.28
SR16a	3.23 ± 0.28	0.87 ± 0.08	13.30 ± 1.02	1.37 ± 0.21	18.77 ± 1.08
SR16b	0.79 ± 0.13	0.34 ± 0.05	0.70 ± 0.23	0.24 ± 0.06	2.07 ± 0.28

Table B.5: Simulated yields for 35.9 fb^{-1} of the rare SM backgrounds in the 23 off-Z signal regions and the associated statistical uncertainty from the size of the samples. Shown uncertainty is MC statistics.

SR1a	9.78 ± 0.47	16.65 ± 0.57	49.35 ± 1.95	18.68 ± 3.06	94.47 ± 3.70
SR1b	0.14 ± 0.05	0.25 ± 0.06	0.30 ± 0.15	0.05 ± 0.06	0.74 ± 0.18
SR2a	1.26 ± 0.19	3.90 ± 0.26	5.47 ± 0.64	1.67 ± 0.38	12.32 ± 0.81
SR2b	0.14 ± 0.06	0.19 ± 0.04	0.07 ± 0.07	0.14 ± 0.06	0.54 ± 0.12
SR3a	1.60 ± 0.21	1.80 ± 0.17	4.47 ± 0.59	2.45 ± 0.92	10.32 ± 1.13
SR3b	0.05 ± 0.04	0.03 ± 0.01	0.10 ± 0.08	0.01 ± 0.01	0.19 ± 0.09
SR4a	0.62 ± 0.13	1.21 ± 0.13	2.00 ± 0.40	0.69 ± 0.19	4.52 ± 0.48
SR4b	0.02 ± 0.02	0.05 ± 0.01	0.05 ± 0.05	0.07 ± 0.07	0.19 ± 0.09
SR5	17.64 ± 0.62	32.33 ± 0.77	3.99 ± 0.56	3.37 ± 1.01	57.33 ± 1.52
SR6	3.24 ± 0.26	6.81 ± 0.31	0.65 ± 0.23	0.05 ± 0.13	10.75 ± 0.49
SR7	3.24 ± 0.29	4.62 ± 0.25	0.62 ± 0.22	0.38 ± 0.29	8.85 ± 0.53
SR8	1.27 ± 0.17	2.48 ± 0.18	0.39 ± 0.17	0.06 ± 0.04	4.20 ± 0.30
SR9	8.84 ± 0.40	15.46 ± 0.48	0.14 ± 0.10	0.01 ± 0.06	24.45 ± 0.64
SR10	1.56 ± 0.16	2.94 ± 0.18	0.00 ± 0.00	0.00 ± 0.00	4.49 ± 0.25
SR11	1.41 ± 0.23	2.58 ± 0.20	0.00 ± 0.00	0.01 ± 0.01	4.00 ± 0.30
SR12	0.67 ± 0.12	1.21 ± 0.10	0.07 ± 0.07	0.00 ± 0.00	1.95 ± 0.17
SR13	0.66 ± 0.13	1.44 ± 0.15	0.00 ± 0.00	0.01 ± 0.01	2.11 ± 0.20
SR14a	3.07 ± 0.33	4.88 ± 0.26	1.60 ± 0.35	0.83 ± 0.49	10.38 ± 0.73
SR14b	0.01 ± 0.06	0.08 ± 0.05	0.16 ± 0.11	0.06 ± 0.04	0.28 ± 0.14
SR15a	1.65 ± 0.23	3.77 ± 0.22	1.56 ± 0.35	0.41 ± 0.16	7.39 ± 0.50
SR15b	0.03 ± 0.05	0.21 ± 0.05	0.08 ± 0.08	0.08 ± 0.05	0.40 ± 0.12
SR16a	1.00 ± 0.16	2.80 ± 0.20	1.48 ± 0.34	0.92 ± 0.18	6.19 ± 0.46
SR16b	0.13 ± 0.05	0.19 ± 0.04	0.00 ± 0.00	0.06 ± 0.04	0.37 ± 0.08

C

MEASUREMENT OF THE TOP PAIR PRODUCTION IN ASSOCIATION WITH A W BOSON

C.1 LIST OF SM SIGNAL AND BACKGROUND SIMULATIONS

Table C.1 lists all MC simulations for SM signal and background processes used in the $t\bar{t}W$ analysis. The cross sections used for the normalizations of the background MC samples are calculated from the matrix element of the MC generator used. The NNPDF3.0 [120, 137] PDFs are used for the hard scattering process in the samples. The modeling of the underlying event is generated using the CUETP8M1 [131, 138] for most of the samples, apart from the $t\bar{t}H$ and ttVV samples where the CUETP8M2 and CUETP8M2T4 tunes [139] are utilized.

The background simulations listed in the 'nonprompt' and 'charge mis-ID' categories are not used for the final result of the analysis where this type of background is estimated with a tight-to-loose technique. Instead, these simulations are used for the closure tests to validate the nonprompt lepton background estimation and to measure charge mis-ID in the MC simulation and validation in data.

Category	Sample Name	Generator	Matching scheme	σ pb
Signal	$t\bar{t}W$, W to $1\ell 1\nu$	MADGRAPH5_aMC@NLO	FxFx	0.2043
nonprompt	$t\bar{t}$ to 1ℓ	Madgraph	MLM	364.36
charge mis-ID	$t\bar{t}$ to 2ℓ	Madgraph	MLM	87.315
	DY to 2ℓ ($m_{\ell\ell} > 50$ GeV)	MADGRAPH5_aMC@NLO	FxFx	6024
WZ	WZ to $3\ell 1\nu$	MADGRAPH5_aMC@NLO	FxFx	4.4297
$t\bar{t}X$	$t\bar{t}H$, H to all, except $b\bar{b}$	Powheg	-	0.2151
	$t\bar{t}Z$, Z to 2ℓ or 2ν ($m_{\ell\ell} > 10$ GeV)	MADGRAPH5_aMC@NLO	-	0.2529
	$t\bar{t}Z$, Z to 2ℓ (1 GeV $< m_{\ell\ell} <$ 10 GeV)	Madgraph	MLM	0.0283
	tZq, Z to 2ℓ	MADGRAPH5_aMC@NLO	-	0.0758
	tHq, H inclusive	Madgraph	-	0.261
	tHW, H inclusive	Madgraph	-	0.01561
	tWZ, Z to 2ℓ	Madgraph	-	0.01123
	$t\bar{t}t\bar{t}$	MADGRAPH5_aMC@NLO	-	0.009103
	$t\bar{t}\gamma$	MADGRAPH5_aMC@NLO	FxFx	3.697
	$t\gamma$	MADGRAPH5_aMC@NLO	-	2.967
	$t\bar{t}WW$	Madgraph	-	0.007829
	$t\bar{t}WZ$	Madgraph	-	0.002919
	$t\bar{t}ZZ$	Madgraph	-	0.001573
Rare SM	ZZ, both Z to 2ℓ	Powheg	-	1.256
	HZZ, both Z to 2ℓ	Powheg	-	0.0121
	VH, H to all, except $b\bar{b}$	MADGRAPH5_aMC@NLO	FxFx	0.9561
	$W^{\pm}W^{\pm}$	Madgraph	-	0.03711
	ZZZ	MADGRAPH5_aMC@NLO	-	0.01398
	WZZ	MADGRAPH5_aMC@NLO	-	0.05565
	WWZ	MADGRAPH5_aMC@NLO	-	0.1651
	WWW	MADGRAPH5_aMC@NLO	-	0.2086
	$Z\gamma$	MADGRAPH5_aMC@NLO	FxFx	131.3
	$W\gamma$	MADGRAPH5_aMC@NLO	MLM	585.5

Table C.1: List of MC simulations of SM background processes used in the $t\bar{t}W$ analysis. Category name, sample name, generator name, the parton matching scheme and processes cross section are presented.

D

MEASUREMENT OF THE TOP PAIR PRODUCTION IN ASSOCIATION WITH A Z BOSON

d.1 LIST OF SM SIGNAL AND BACKGROUND SIMULATIONS

Table C.1 lists all MC simulations for SM signal and background processes used in the $t\bar{t}Z$ analysis. The cross sections used for the normalizations of the background MC samples are calculated from the ME of the MC generator used. The NNPDF3.0 (NNPDF3.1) [120, 137] PDFs are used for the hard scattering process in 2016 (2017) samples. The modeling of the underlying event is generated using the CUETP8M1 [131, 138] and CP5 tunes [140] for simulated samples corresponding to the 2016 and 2017 data sets, respectively. In 2016, the CUETP8M2 and CUETP8M2T4 tunes [139] are used for the $t\bar{t}H$ and ttVV samples.

The background simulations listed in the 'nonprompt' category are not used for the final result of the analysis where this type of background is estimated with a tight-to-loose technique. Instead, these simulations are used for the closure tests to validate the nonprompt lepton background estimation in the MC simulation and validation in data.

Category	Sample Name	Generator	Matching scheme	σ pb
Signal	$t\bar{t}Z$, Z to 2ℓ or 2ν ($m_{\ell\ell} > 10$ GeV)	MADGRAPH5_aMC@NLO	-	0.2529
nonprompt	$t\bar{t}$ to 2ℓ	Madgraph	MLM	87.315
	$t\bar{t}$ to 1ℓ	Madgraph	MLM	364.36
	DY to 2ℓ ($m_{\ell\ell} > 50$ GeV)	MADGRAPH5_aMC@NLO	FxFx	6024
	DY to 2ℓ (10 GeV $< m_{\ell\ell} < 50$ GeV)	MADGRAPH5_aMC@NLO	FxFx	18610
WZ	WZ to $3\ell 1\nu$	MADGRAPH5_aMC@NLO	FxFx	4.4297
ZZ	ZZ, both Z to 2ℓ	Powheg	-	1.256
	HZZ, both Z to 2ℓ	Powheg	-	0.0121
	gg→ZZ, one Z to 2ℓ, one Z to 2τ	MCFM [141]	-	0.005423
	gg→ZZ, both Z to 2ℓ	MCFM [141]	-	0.0027
$t\bar{t}X$	$t\bar{t}H$, H to all, except $b\bar{b}$	Powheg	-	0.2151
	$t\bar{t}Z$, Z to 2ℓ (1 GeV $< m_{\ell\ell} < 10$ GeV)	Madgraph	MLM	0.0283
	tZq, Z to 2ℓ	MADGRAPH5_aMC@NLO	-	0.0758
	$t\bar{t}W$, W to $1\ell 1\nu$	MADGRAPH5_aMC@NLO	FxFx	0.2043
	tHq, H inclusive	Madgraph	-	0.261
	tHW, H inclusive	Madgraph	-	0.01561
	tWZ, Z to 2ℓ	Madgraph	-	0.01123
	$t\bar{t}t\bar{t}$	MADGRAPH5_aMC@NLO	-	0.009103
	$t\bar{t}WW$	Madgraph	-	0.007829
	$t\bar{t}WZ$	Madgraph	-	0.002919
	$t\bar{t}ZZ$	Madgraph	-	0.001573
Xγ	$t\bar{t}\gamma$	MADGRAPH5_aMC@NLO	FxFx	3.697
	$t\gamma$	MADGRAPH5_aMC@NLO	-	2.967
	Zγ	MADGRAPH5_aMC@NLO	FxFx	131.3
	Wγ	MADGRAPH5_aMC@NLO	MLM	585.5
Rare SM	ZZZ	MADGRAPH5_aMC@NLO	-	0.01398
	WZZ	MADGRAPH5_aMC@NLO	-	0.05565
	WWZ	MADGRAPH5_aMC@NLO	-	0.1651
	WWW	MADGRAPH5_aMC@NLO	-	0.2086

Table D.1: List of MC simulations of SM background processes used in the $t\bar{t}Z$ analysis. Category name, sample name, generator name, the parton matching scheme and processes cross section are presented.

d.2 HIGH-LEVEL TRIGGER PATHS

Table D.2, are used to select QCD multijet events for measuring the lepton misidentification rate, which is used for the estimation of the nonprompt lepton background. The auxiliary triggers have been pre-scaled to reduce their rate.

HLT_Mu3_PFJet40	
HLT_Mu8	HLT_Ele8_CaloIdM_TrackIdM_PFJet30
HLT_Mu17	HLT_Ele17_CaloIdM_TrackIdM_PFJet30
HLT_Mu27	HLT_Ele23_CaloIdM_TrackIdM_PFJet30

Table D.2: List of auxiliary triggers used for the misidentification rate measurement in the $t\bar{t}Z$ analysis

d.3 VALIDATION OF THE TIGHT-TO-LOOSE METHOD IN MC

The closure test is performed for the tight-to-loose method and is validated in $t\bar{t}$ (Fig. D.1) and DY (Fig. D.2) MC simulation events that pass selection described in Section 8.2.

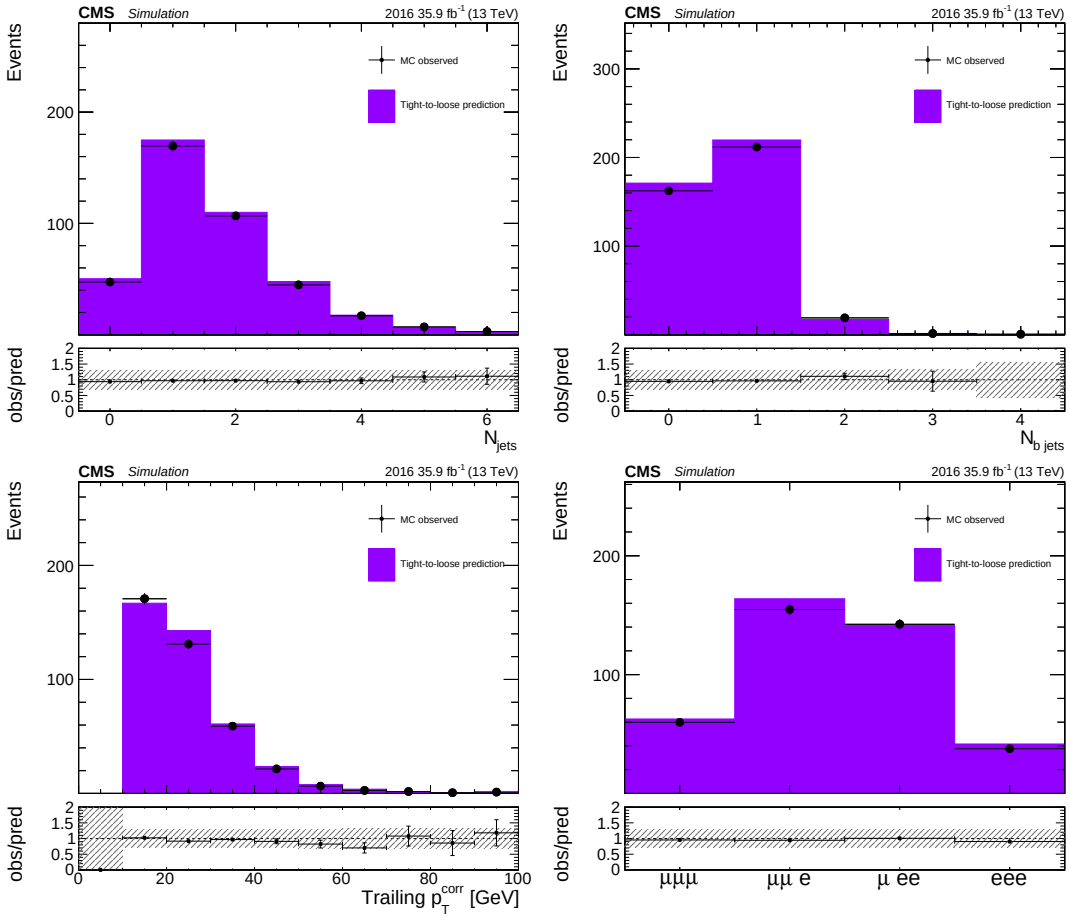


Figure D.1: Validation for the nonprompt lepton background estimation technique in simulation. The distributions of N_{j}, N_{b} and trailing lepton p_{T} and the predicted number of events as a function of the flavor composition as obtained from a $t\bar{t}$ sample are shown.

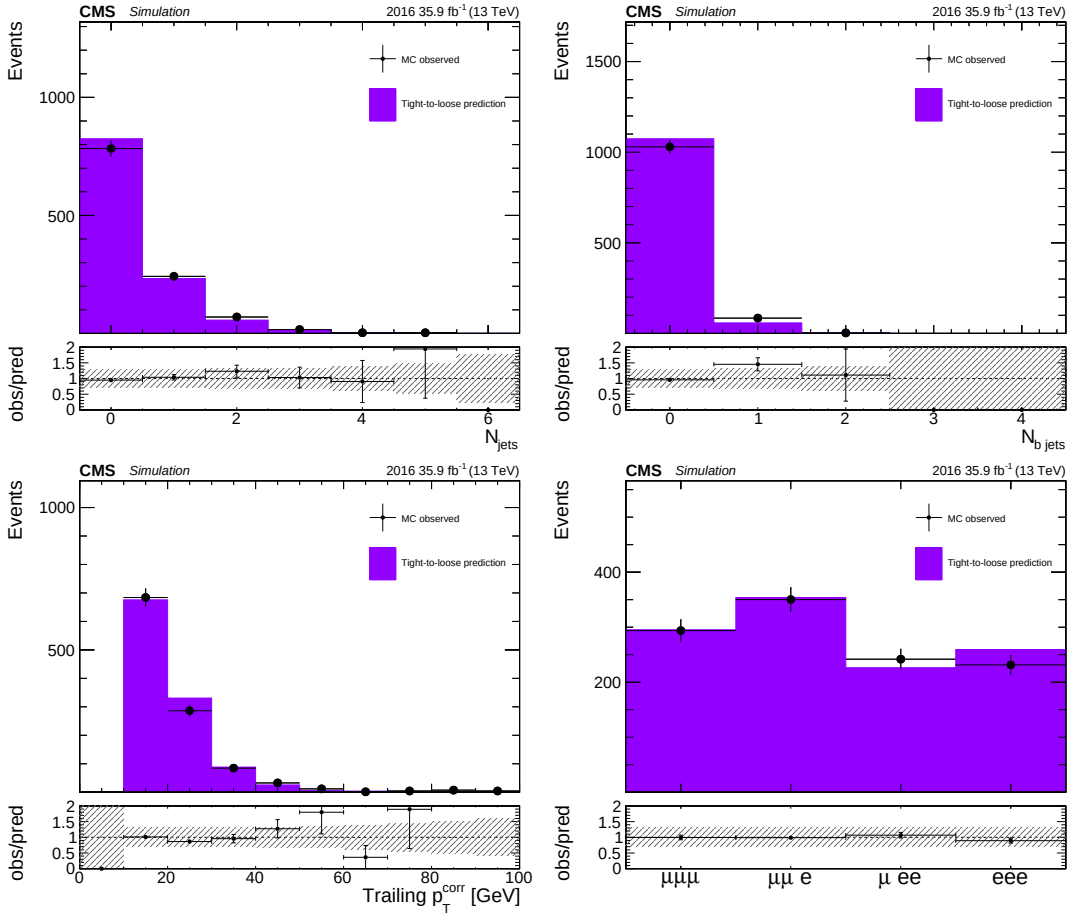


Figure D.2: Validation for the nonprompt lepton background estimation technique in simulation. The distributions of N_{j}, N_{b} and trailing lepton p_{T} and the predicted number of events as a function of the flavor composition as obtained from a DY sample are shown.

d.4 EXPERIMENTAL SYSTEMATIC UNCERTAINTIES

The magnitude and shape of all uncertainties affecting the yield of the processes in $t\bar{t}X$ category and the WZ process are shown in Fig. D.3 and D.4 in the signal regions used for $t\bar{t}Z$ cross section and signal strength modifier extraction.

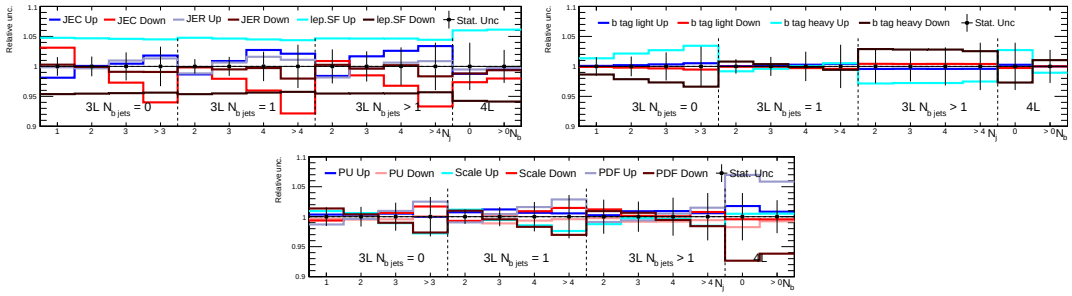


Figure D.3: Input systematic uncertainties for processes in $t\bar{t}X$ category in each of 14 signal regions used for signal strength modifier extraction. Uncertainties for JEC, JER, lepton scale factors, heavy and light flavor b tagging scale factors, PU, μ_R and μ_F scales, PDF and modeling of the parton shower are shown.

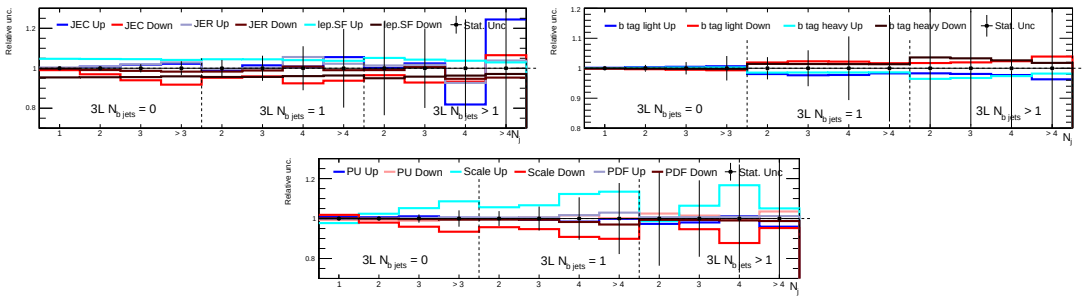


Figure D.4: Input systematic uncertainties for WZ process in each of 14 signal regions used for signal strength modifier extraction. Uncertainties for JEC, JER, lepton scale factors, heavy and light flavor b tagging scale factors, PU, μ_R and μ_F scales, PDF and modeling of the parton shower are shown.

d.5 IMPACT PARAMETER PLOTS

The pull and constrain of each nuisance parameter as obtained from the maximization of the Likelihood function are shown in impact plots in Figures D.5, D.6 and D.7. Individual impact plots are shown for the fit in 2016 and 2017 data sets, and as well for a combination of the two. The corresponding name in the impact plot for a particular nuisance can be found in Table. D.3.

Source	Name in the impact plot
Integrated luminosity	lumi
PU modeling	pileup
Trigger eff.	trigger(2016,2017)
Lepton ID efficiency	lepSFsyst, lepSFstat(2016,2017), lepSFReco
Jet energy scale	jec
Jet energy resolution	jer
B tagging light fl.	bTag_udsg(2016,2017)
B tagging heavy fl	bTag_bc(2016,2017)
Choice in μ_R and μ_F	scaleAcc
PDF choice	pdfAcc
Parton shower	ISRandFSR
WZ cross section	WZ, WZcs(for additional 20% in high N_j categories)
WZ gluon splitting	WZbb
ZZ cross section	ZZ
$t\bar{t}X$ background	ttX
Xγ background	Xgamma
Nonprompt background	nonprompt
Rare SM background	rare
Stat. unc.	prop_binbin{1-2}_bin{0-13}

Table D.3: The table links the names from the table with systematics uncertainties to the names in the impact plots.

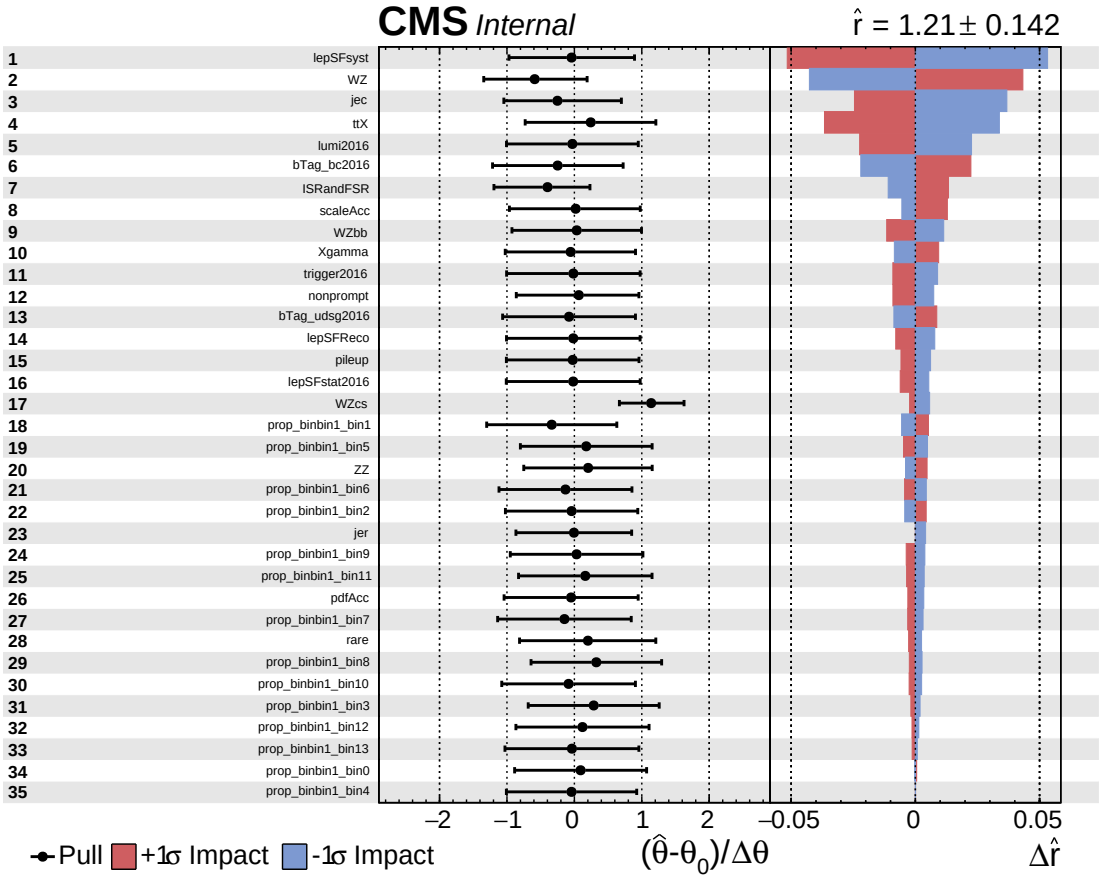


Figure D.5: Impact, pull and constrain of individual nuisances on measured signal strength in three- and four-lepton signal regions in the 2016 data set.

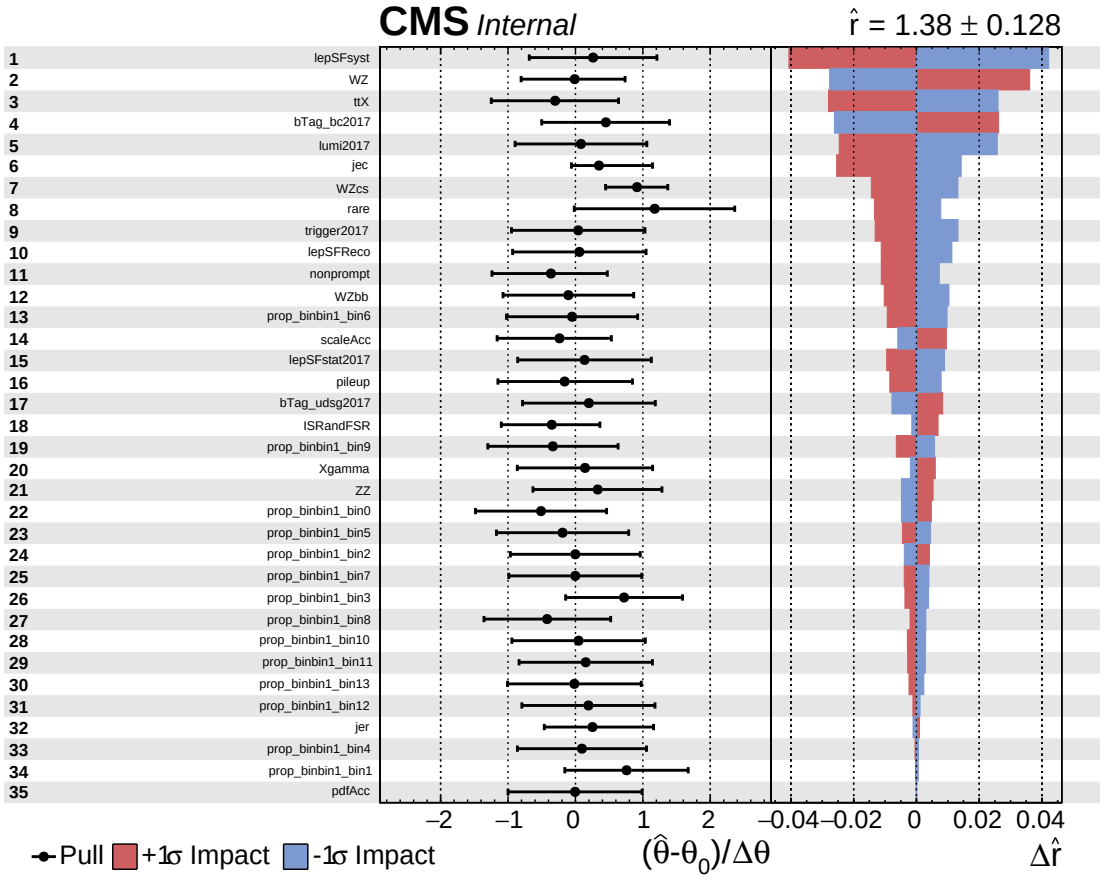


Figure D.6: IImpact, pull and constrain of individual nuisances on measured signal strength in three- and four-lepton signal regions in the 2017 data set.

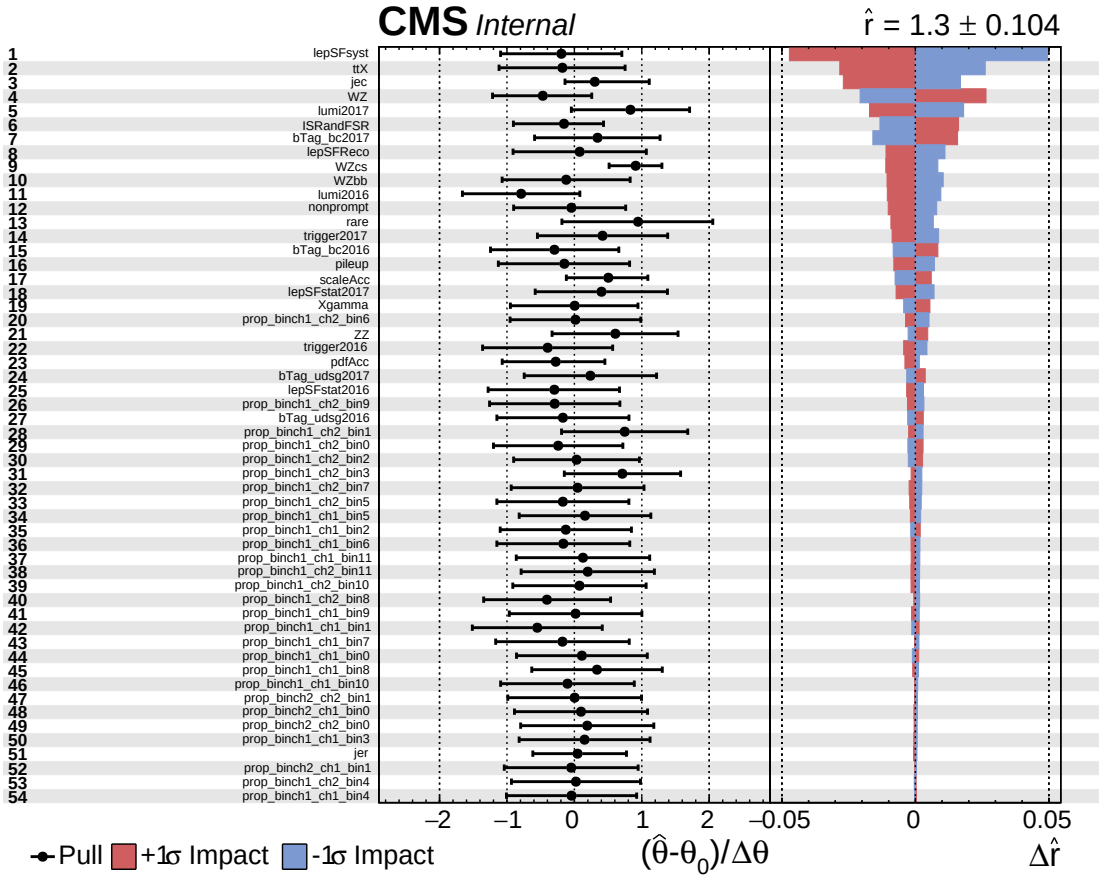


Figure D.7: Impact, pull and constrain of individual nuisances on measured signal strength in three- and four-lepton signal regions in the 2016 and 2017 data sets.

d.6 KINEMATICAL VARIABLE DISTRIBUTIONS

Kinematical variable distributions in the most significant and dominated $t\bar{t}Z$ regions are shown in Fig. D.8 and D.9. These regions are defined as following: in the three-lepton final state at least 3 jets and 1 b-tagged jet and in four-lepton final state at least 2 jets and 1 b tagged jet are required in the event selection. In four lepton final state for non-Z lepton the one with highest lepton p_T was picked.

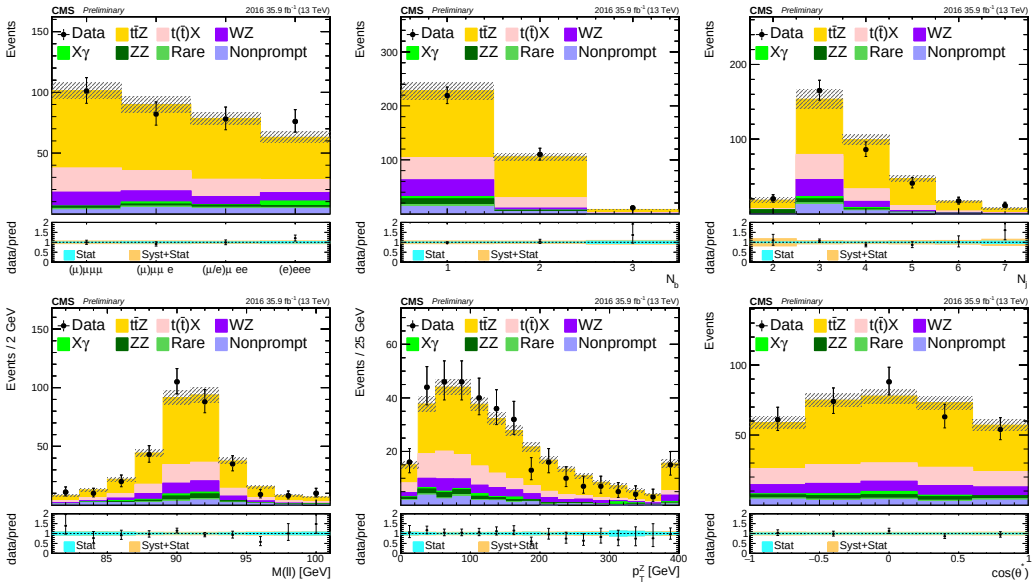


Figure D.8: Observed data in a $t\bar{t}Z$ dominated region, compared to signal and background yields, as obtained from the fit only with the 2016 data set. Event distributions are shown as a function of the lepton flavor (upper left), N_b (upper middle), N_j (upper right), dilepton invariant mass of OCSF pair (lower left), p_T(Z) (lower middle), and $\cos\theta^*_Z$ (lower right). The hatched band shows the total uncertainty associated with the signal and background predictions, as obtained from the fit.

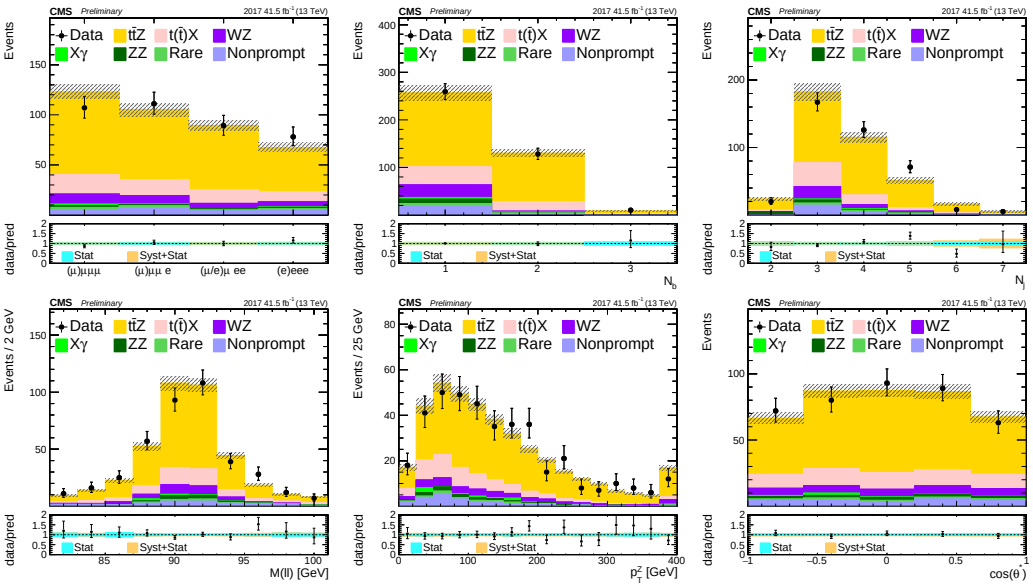


Figure D.9: Observed data in a $t\bar{t}Z$ dominated region, compared to signal and background yields, as obtained from the fit only with the 2017 data set. Event distributions are shown as a function of the lepton flavor (upper left), N_b (upper middle), N_j (upper right), dilepton invariant mass of OCSF pair (lower left), $p_T(Z)$ (lower middle), and $\cos\theta^*_Z$ (lower right). The hatched band shows the total uncertainty associated with the signal and background predictions, as obtained from the fit.

BIBLIOGRAPHY

[1] Serguei Chatrchyan et al. "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC." In: *Phys. Lett.* B716 (2012), pp. 30–61. DOI: 10.1016/j.physletb.2012.08.021. arXiv: 1207.7235 [hep-ex] (cit. on pp. 1, 16, 17).

[2] Georges Aad et al. "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC." In: *Phys. Lett.* B716 (2012), pp. 1–29. DOI: 10.1016/j.physletb.2012.08.020. arXiv: 1207.7214 [hep-ex] (cit. on pp. 1, 16, 17).

[3] Albert M Sirunyan et al. "Observation of single top quark production in association with a Z boson in proton-proton collisions at $\sqrt{s} = 13$ TeV." In: *Submitted to: Phys. Rev. Lett.* (2018). arXiv: 1812.05900 [hep-ex] (cit. on pp. 1, 217, 225).

[4] Michael E. Peskin and Daniel V. Schroeder. *An Introduction to quantum field theory*. Reading, USA: Addison-Wesley, 1995. ISBN: 9780201503975, 0201503972. URL: http://www.slac.stanford.edu/~mpeskin/QFT.html (cit. on pp. 5, 8, 9, 12).

[5] David Griffiths. *Introduction to elementary particles*. 2008. ISBN: 9783527406012 (cit. on pp. 5, 15).

[6] S. N. Bose. "Planck's law and light quantum hypothesis." In: *Z. Phys.* 26 (1924), pp. 178–181. DOI: 10.1007/BF01327326 (cit. on p. 5).

[7] Paul A. M. Dirac. "On the Theory of quantum mechanics." In: *Proc. Roy. Soc. Lond.* A112 (1926), pp. 661–677. DOI: 10.1098/rspa.1926.0133 (cit. on p. 5).

[8] Paul Adrien Maurice Dirac. *The Principles of Quantum Mechanics; 3rd ed.* International series of monographs on physics. Oxford: Clarendon Press, 1947. URL: https://cds.cern.ch/record/107130 (cit. on pp. 5, 20, 21).

[9] D. H. Perkins. *Introduction to high energy physics*. 1982. ISBN: 9780521621960 (cit. on pp. 5, 9).

[10] C. Patrignani et al. "Review of Particle Physics." In: *Chin. Phys.* C40.10 (2016), p. 100001. DOI: 10.1088/1674-1137/40/10/100001 (cit. on pp. 6, 7, 61, 174, 218).

[11] D. Decamp et al. "Determination of the Number of Light Neutrino Species." In: *Phys. Lett.* B231 (1989), pp. 519–529. DOI: 10.1016/0370-2693(89)90704-1 (cit. on p. 6).

[12] G. Arnison et al. "Experimental Observation of Isolated Large Transverse Energy Electrons with Associated Missing Energy at s**(1/2) = 540-GeV." In: *Phys. Lett.* B122 (1983). [,611(1983)], pp. 103–116. DOI: 10.1016/0370-2693(83)91177-2 (cit. on p. 6).

[13] M. Banner et al. "Observation of Single Isolated Electrons of High Transverse Momentum in Events with Missing Transverse Energy at the CERN anti-p p Collider." In: *Phys. Lett.* B122 (1983). [,7.45(1983)], pp. 476–485. DOI: 10.1016/0370-2693(83)91605-2 (cit. on p. 6).

[14] David Galbraith. *UX: Standard Model of the Standard Model*. http://davidgalbraith.org/portfolio/ux-standard-model-of-the-standard-model/http://www.texample.net/tikz/examples/model-physics/ (cit. on p. 7).

[15] L. D. Landau and E. M. Lifschits. *The Classical Theory of Fields*. Vol. Volume 2. Course of Theoretical Physics. Oxford: Pergamon Press, 1975. ISBN: 9780080181769 (cit. on p. 6).

[16] Emmy Noether. "Invariant Variation Problems." In: *Gott. Nachr.* 1918 (1918). [Transp. Theory Statist. Phys.1,186(1971)], pp. 235–257. DOI: 10.1080/00411457108231446. arXiv: physics/0503066 [physics] (cit. on p. 7).

[17] Peter Woit. *Quantum Theory, Groups and Representations*. Springer, 2017. ISBN: 9783319646107, 9783319646121. DOI: 10.1007/978-3-319-64612-1. URL: http://inference-review.com/article/woits-way (cit. on pp. 8, 9).

[18] Alessandro De Angelis and Mário João Martins Pimenta. *Introduction to particle and astroparticle physics: questions to the Universe*. Undergraduate lecture notes in physics. Milan: Springer, 2013. URL: http://cds.cern.ch/record/1529338 (cit. on pp. 9, 11, 12).

[19] R. P. Feynman. "The Theory of positrons." In: *Phys. Rev.* 76 (1949). [,88(1949)], pp. 749–759. DOI: 10.1103/PhysRev.76.749 (cit. on pp. 9, 10, 38).

[20] Salvatore Mele. "Measurement of the running of the electromagnetic coupling at LEP." In: *PoS* HEP2005 (2006), p. 286. DOI: 10.22323/1.021.0286. arXiv: hep-ex/0601045 [hep-ex] (cit. on p. 11).

[21] Morton Hamermesh. *Group theory and its application to physical problems*. New York, NY: Dover, 1989. URL: https://cds.cern.ch/record/1123140 (cit. on p. 11).

[22] S. L. Glashow. "Partial Symmetries of Weak Interactions." In: *Nucl. Phys.* 22 (1961), pp. 579–588. DOI: 10.1016/0029-5582(61)90469-2 (cit. on p. 13).

[23] Abdus Salam. "Weak and Electromagnetic Interactions." In: *Conf. Proc.* C680519 (1968), pp. 367–377 (cit. on p. 13).

[24] Steven Weinberg. "A Model of Leptons." In: *Phys. Rev. Lett.* 19 (1967), pp. 1264–1266. DOI: 10.1103/PhysRevLett.19.1264 (cit. on p. 13).

[25] T. Nakano and K. Nishijima. "Charge Independence for V-particles." In: *Prog. Theor. Phys.* 10 (1953), pp. 581–582. DOI: 10.1143/PTP.10.581 (cit. on p. 14).

[26] Kazuhiko Nishijima. "Charge Independence Theory of V Particles." In: *Prog. Theor. Phys.* 13.3 (1955), pp. 285–304. DOI: 10.1143/PTP.13.285 (cit. on p. 14).

[27] M. Gell-Mann. "The interpretation of the new particles as displaced charge multiplets." In: *Nuovo Cim.* 4.S2 (1956), pp. 848–866. DOI: 10.1007/BF02748000 (cit. on p. 14).

[28] F. Englert and R. Brout. "Broken Symmetry and the Mass of Gauge Vector Mesons." In: *Phys. Rev. Lett.* 13 (1964). [,157(1964)], pp. 321–323. DOI: 10.1103/PhysRevLett.13.321 (cit. on p. 15).

[29] Peter W. Higgs. "Broken Symmetries and the Masses of Gauge Bosons." In: *Phys. Rev. Lett.* 13 (1964). [,160(1964)], pp. 508–509. DOI: 10.1103/PhysRevLett.13.508 (cit. on p. 15).

[30] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble. "Global Conservation Laws and Massless Particles." In: *Phys. Rev. Lett.* 13 (1964). [,162(1964)], pp. 585–587. DOI: 10.1103/PhysRevLett.13.585 (cit. on p. 15).

[31] Albert M Sirunyan et al. "Measurements of properties of the Higgs boson decaying into the four-lepton final state in pp collisions at $\sqrt{s} = 13$ TeV." In: *JHEP* 11 (2017), p. 047. DOI: 10.1007/JHEP11(2017)047. arXiv: 1706.09936 [hep-ex] (cit. on p. 16).

[32] Albert M Sirunyan et al. "Measurements of the Higgs boson width and anomalous HVV couplings from on-shell and off-shell production in the four-lepton final state." In: (2019). arXiv: 1901.00174 [hep-ex] (cit. on p. 16).

[33] *Measurements of properties of the Higgs boson in the four-lepton final state in proton-proton collisions at* $\sqrt{s} = 13$ TeV. Tech. rep. CMS-PAS-HIG-19-001. Geneva: CERN, 2019. URL: https://cds.cern.ch/record/2668684 (cit. on p. 16).

[34] Albert M Sirunyan et al. "Observation of the Higgs boson decay to a pair of τ leptons with the CMS detector." In: *Phys. Lett.* B779 (2018), pp. 283–316. DOI: 10.1016/j.physletb.2018.02.004. arXiv: 1708.00373 [hep-ex] (cit. on p. 16).

[35] A. M. Sirunyan et al. "Observation of Higgs boson decay to bottom quarks." In: *Phys. Rev. Lett.* 121.12 (2018), p. 121801. DOI: 10.1103/PhysRevLett.121.121801. arXiv: 1808.08242 [hep-ex] (cit. on p. 16).

[36] Albert M Sirunyan et al. "Observation of $t\bar{t}H$ production." In: *Phys. Rev. Lett.* 120.23 (2018), p. 231801. DOI: 10.1103/PhysRevLett.120.231801. arXiv: 1804.02610 [hep-ex] (cit. on pp. 16, 51, 52).

[37] Albert M. Sirunyan et al. "Search for the Higgs boson decaying to two muons in proton-proton collisions at $\sqrt{s} = 13$ TeV." In: *Phys. Rev. Lett.* 122.2 (2019), p. 021801. DOI: 10.1103/PhysRevLett.122.021801. arXiv: 1807.06325 [hep-ex] (cit. on p. 16).

[38] Albert M Sirunyan et al. "Combined measurements of Higgs boson couplings in proton-proton collisions at $\sqrt{s} = 13$ TeV." In: *Submitted to: Eur. Phys. J.* (2018). arXiv: 1809.10733 [hep-ex] (cit. on p. 17).

[39] P. A. R. Ade et al. "Planck 2013 results. I. Overview of products and scientific results." In: *Astron. Astrophys.* 571 (2014), A1. DOI: 10.1051/0004-6361/201321529. arXiv: 1303.5062 [astro-ph.CO] (cit. on p. 17).

[40] Edvige Corbelli and Paolo Salucci. "The Extended Rotation Curve and the Dark Matter Halo of M33." In: *Mon. Not. Roy. Astron. Soc.* 311 (2000), pp. 441–447. DOI: 10.1046/j.1365-8711.2000.03075.x. arXiv: astro-ph/9909252 [astro-ph] (cit. on p. 17).

[41] Laurent Canetti, Marco Drewes, and Mikhail Shaposhnikov. "Matter and Antimatter in the Universe." In: *New J. Phys.* 14 (2012), p. 095012. DOI: 10.1088/1367-2630/14/9/095012. arXiv: 1204.4186 [hep-ph] (cit. on p. 18).

[42] F. Abe et al. "Observation of top quark production in $\bar{p}p$ collisions." In: *Phys. Rev. Lett.* 74 (1995), pp. 2626–2631. DOI: 10.1103/PhysRevLett.74.2626. arXiv: hep-ex/9503002 [hep-ex] (cit. on p. 18).

[43] S. Abachi et al. "Search for high mass top quark production in $p\bar{p}$ collisions at $\sqrt{s} = 1.8$ TeV." In: *Phys. Rev. Lett.* 74 (1995), pp. 2422–2426. DOI: 10.1103/PhysRevLett.74.2422. arXiv: hep-ex/9411001 [hep-ex] (cit. on p. 18).

[44] Raoul Röntsch and Markus Schulze. "Constraining the top-Z coupling through $t\bar{t}Z$ production at the LHC." In: *Nucl. Part. Phys. Proc.* 273-275 (2016), pp. 2311–2316. DOI: 10.1016/j.nuclphysbps.2015.09.377 (cit. on p. 19).

[45] Albert M Sirunyan et al. "Measurement of the cross section for top quark pair production in association with a W or Z boson in proton-proton collisions at $\sqrt{s} = 13$ TeV." In: *JHEP* 08 (2018), p. 011. DOI: 10.1007/JHEP08(2018)011. arXiv: 1711.02547 [hep-ex] (cit. on pp. 19, 20, 52, 177, 207, 218).

[46] Morad Aaboud et al. "Combination of inclusive and differential top-antitop charge asymmetry measurements using ATLAS and CMS data at center-of-mass energy of 7 and 8 TeV." In: *JHEP* 04 (2018), p. 033. DOI: 10.1007/JHEP04(2018)033. arXiv: 1709.05327 [hep-ex] (cit. on p. 19).

[47] F. Maltoni, M. L. Mangano, I. Tsinikos, and M. Zaro. "Top-quark charge asymmetry and polarization in ttW production at the LHC." In: *Phys. Lett.* B736 (2014), pp. 252–260. DOI: 10.1016/j.physletb.2014.07.033. arXiv: 1406.3262 [hep-ph] (cit. on pp. 19, 224).

[48] Kenneth G. Wilson. "Nonlagrangian models of current algebra." In: *Phys. Rev.* 179 (1969), pp. 1499–1512. DOI: 10.1103/PhysRev.179.1499 (cit. on p. 19).

[49] Anna Elizabeth Woodard and Kevin Lannon. "Effective field theory interpretation for measurements of top quark pair-production in association with a W or Z boson." Presented 27 Feb 2018. 2018. URL: http://cds.cern.ch/record/2649938 (cit. on p. 20).

[50] Stephen P. Martin. "A Supersymmetry primer." In: (1997). [Adv. Ser. Direct. High Energy Phys.18,1(1998)], pp. 1–98. DOI: 10.1142/9789812839657_0001,10.1142/9789814307505_0001. arXiv: hep-ph/9709356 [hep-ph] (cit. on pp. 20–22).

[51] Borut Bajc, Junji Hisano, Takumi Kuwahara, and Yuji Omura. "Threshold corrections to dimension-six proton decay operators in non-minimal SUSY SU (5) GUTs." In: *Nucl. Phys.* B910 (2016), pp. 1–22. DOI: 10.1016/j.nuclphysb.2016.06.017. arXiv: 1603.03568 [hep-ph] (cit. on p. 20).

[52] *Summary of comparison plots in simplified models spectra for the 13TeV dataset.* https://twiki.cern.ch/twiki/bin/view/CMSPublic/SUSYSMSSummaryPlots13TeV (cit. on p. 23).

[53] Daniele Alves. "Simplified Models for LHC New Physics Searches." In: *J. Phys.* G39 (2012). Ed. by Nima Arkani-Hamed, Sanjay Arora, Yang Bai, Matthew Baumgart, Joshua Berger, Matthew Buckley, Bart Butler, Spencer Chang, Hsin-Chia Cheng, Clifford Cheung, et al., p. 105005. DOI: 10.1088/0954-3899/39/10/105005. arXiv: 1105.2838 [hep-ph] (cit. on p. 24).

[54] Johan Alwall, Philip Schuster, and Natalia Toro. "Simplified Models for a First Characterization of New Physics at the LHC." In: *Phys. Rev.* D79 (2009), p. 075020. DOI: 10.1103/PhysRevD.79.075020. arXiv: 0810.3921 [hep-ph] (cit. on p. 24).

[55] Serguei Chatrchyan et al. "Interpretation of Searches for Supersymmetry with simplified Models." In: *Phys. Rev.* D88.5 (2013), p. 052017. DOI: 10.1103/PhysRevD.88.052017. arXiv: 1301.2175 [hep-ex] (cit. on p. 24).

[56] Oliver Sim Brüning, Paul Collier, P Lebrun, Stephen Myers, Ranko Ostojic, John Poole, and Paul Proudlock. *LHC Design Report*. CERN Yellow Reports: Monographs. Geneva: CERN, 2004. URL: https://cds.cern.ch/record/782076 (cit. on p. 25).

[57] S. Chatrchyan et al. "The CMS Experiment at the CERN LHC." In: *JINST* 3 (2008), S08004. DOI: 10.1088/1748-0221/3/08/S08004 (cit. on pp. 25, 30, 33).

[58] D. Brandt, H. Burkhardt, M. Lamont, S. Myers, and J. Wenninger. "Accelerator physics at LEP." In: *Rept. Prog. Phys.* 63 (2000), pp. 939–1000. DOI: 10.1088/0034-4885/63/6/203 (cit. on p. 26).

[59] Esma Mobs. "The CERN accelerator complex. Complexe des accélérateurs du CERN." In: (2016). General Photo. URL: https://cds.cern.ch/record/2197559 (cit. on p. 27).

[60] S. Schael et al. "Precision electroweak measurements on the Z resonance." In: *Phys. Rept.* 427 (2006), pp. 257–454. DOI: 10.1016/j.physrep.2005.12.006. arXiv: hep-ex/0509008 [hep-ex] (cit. on p. 28).

[61] Tetsuo Abe et al. "Achievements of KEKB." In: *PTEP* 2013 (2013), 03A001. DOI: 10.1093/ptep/pts102 (cit. on p. 28).

[62] *ZEUS Overview*. https://www-zeus.desy.de/~gallo/ZEUSOverview.html (cit. on p. 28).

[63] *Public CMS Luminosity Information*. https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults (cit. on pp. 28, 57, 109).

[64] *CMS Photos, 78 reconstructed vertices in event from high-pileup run 198609*. https://cds.cern.ch/record/1479324 (cit. on p. 28).

[65] *CMS detector design*. http://cms.web.cern.ch/news/cms-detector-design (cit. on p. 29).

[66] Izaak Neutelings. https://wiki.physik.uzh.ch/cms/latex:example_spherical_coordinates. 2017 (cit. on p. 30).

[67] L. Viliani. "CMS tracker performance and readiness for LHC Run II." In: *Nucl. Instrum. Meth.* A824 (2016), pp. 67–69. DOI: 10.1016/j.nima.2015.09.046 (cit. on p. 31).

[68] Serguei Chatrchyan et al. "Description and performance of track and primary-vertex reconstruction with the CMS tracker." In: *JINST* 9.10 (2014), P10009. DOI: 10.1088/1748-0221/9/10/P10009. arXiv: 1405.6569 [physics.ins-det] (cit. on pp. 31, 44).

[69] "CMS: The electromagnetic calorimeter. Technical design report." In: (1997) (cit. on p. 32).

[70] A. Benaglia. "The CMS ECAL performance with examples." In: *JINST* 9 (2014), p. C02008. DOI: 10.1088/1748-0221/9/02/C02008 (cit. on p. 32).

[71] P. Adzic et al. "Energy resolution of the barrel of the CMS electromagnetic calorimeter." In: *JINST* 2 (2007), P04004. DOI: 10.1088/1748-0221/2/04/P04004 (cit. on p. 32).

[72] *XDAQ - CMS Online Software project page*. http://xdaq.web.cern.ch/xdaq/setup/images/HCAL.png (cit. on p. 33).

[73] D. Abbaneo et al. "Studies on the upgrade of the muon system in the forward region of the CMS experiment at LHC with GEMs." In: *JINST* 9.01 (2014), p. C01053. DOI: 10.1088/1748-0221/9/01/C01053 (cit. on p. 34).

[74] A. Colaleo, A. Safonov, A. Sharma, and M. Tytgat. "CMS Technical Design Report for the Muon Endcap GEM Upgrade." In: (2015) (cit. on p. 35).

[75] Vardan Khachatryan et al. "The CMS trigger system." In: *JINST* 12.01 (2017), P01020. DOI: 10.1088/1748-0221/12/01/P01020. arXiv: 1609.02366 [physics.ins-det] (cit. on p. 35).

[76] Jamie Shiers. "The Worldwide LHC Computing Grid (worldwide LCG)." In: *Comput. Phys. Commun.* 177 (2007), pp. 219–223. DOI: 10.1016/j.cpc.2007.02.021 (cit. on p. 35).

[77] David J. Gross and Frank Wilczek. "Ultraviolet Behavior of Nonabelian Gauge Theories." In: *Phys. Rev. Lett.* 30 (1973). [,271(1973)], pp. 1343–1346. DOI: 10.1103/PhysRevLett.30.1343 (cit. on p. 37).

[78] H. David Politzer. "Reliable Perturbative Results for Strong Interactions?" In: *Phys. Rev. Lett.* 30 (1973). [,274(1973)], pp. 1346–1349. DOI: 10.1103/PhysRevLett.30.1346 (cit. on p. 37).

[79] Guido Altarelli and G. Parisi. "Asymptotic Freedom in Parton Language." In: *Nucl. Phys.* B126 (1977), pp. 298–318. DOI: 10.1016/0550-3213(77)90384-4 (cit. on p. 38).

[80] Yuri L. Dokshitzer. "Calculation of the Structure Functions for Deep Inelastic Scattering and e+ e- Annihilation by Perturbation Theory in Quantum Chromodynamics." In: *Sov. Phys. JETP* 46 (1977). [Zh. Eksp. Teor. Fiz.73,1216(1977)], pp. 641–653 (cit. on p. 38).

[81] V. N. Gribov and L. N. Lipatov. "Deep inelastic e p scattering in perturbation theory." In: *Sov. J. Nucl. Phys.* 15 (1972). [Yad. Fiz.15,781(1972)], pp. 438–450 (cit. on p. 38).

[82] Marcello Fanti. *Physics at LHC*. http://www0.mi.infn.it/~fanti/Particelle4/fanti_PhysicsAtLHC.pdf (cit. on p. 38).

[83] Bo Andersson, G. Gustafson, G. Ingelman, and T. Sjostrand. "Parton Fragmentation and String Dynamics." In: *Phys. Rept.* 97 (1983), pp. 31–145. DOI: 10.1016/0370-1573(83)90080-7 (cit. on p. 39).

[84] Torbjörn Sjöstrand, Stefan Ask, Jesper R. Christiansen, Richard Corke, Nishita Desai, Philip Ilten, Stephen Mrenna, Stefan Prestel, Christine O. Rasmussen, and Peter Z. Skands. "An Introduction to PYTHIA 8.2." In: *Comput. Phys. Commun.* 191 (2015), pp. 159–177. DOI: 10.1016/j.cpc.2015.01.024. arXiv: 1410.3012 [hep-ph] (cit. on p. 39).

[85] C. Mena and L. F. Palhares. "Quark-antiquark potentials in non-perturbative models." In: *14th International Workshop on Hadron Physics (Hadron Physics 2018) Florianopolis, Santa Catarina, Brazil, March 18-23, 2018*. 2018. arXiv: 1804.09564 [hep-ph] (cit. on p. 39).

[86] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli, and M. Zaro. "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations." In: *JHEP* 07 (2014), p. 079. DOI: 10.1007/JHEP07(2014)079. arXiv: 1405.0301 [hep-ph] (cit. on p. 39).

[87] Johan Alwall et al. "Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions." In: *Eur. Phys. J.* C53 (2008), pp. 473–500. DOI: 10.1140/epjc/s10052-007-0490-5. arXiv: 0706.2569 [hep-ph] (cit. on p. 39).

[88] Rikkert Frederix and Stefano Frixione. "Merging meets matching in MC@NLO." In: *JHEP* 12 (2012), p. 061. DOI: 10.1007/JHEP12(2012)061. arXiv: 1209.6215 [hep-ph] (cit. on p. 39).

[89] Carlo Oleari. "The POWHEG-BOX." In: *Nucl. Phys. Proc. Suppl.* 205-206 (2010), pp. 36–41. DOI: 10.1016/j.nuclphysbps.2010.08.016. arXiv: 1007.3893 [hep-ph] (cit. on p. 39).

[90] Stefan Höche. "Introduction to parton-shower event generators." In: *Proceedings, Theoretical Advanced Study Institute in Elementary Particle Physics: Journeys Through the Precision Frontier: Amplitudes for Colliders (TASI 2014): Boulder, Colorado, June 2-27, 2014*. 2015, pp. 235–295. DOI: 10.1142/9789814678766_0005. arXiv: 1411.4085 [hep-ph] (cit. on p. 40).

[91] S. Agostinelli et al. "GEANT4: A Simulation toolkit." In: *Nucl. Instrum. Meth.* A506 (2003), pp. 250–303. DOI: 10.1016/S0168-9002(03)01368-8 (cit. on p. 40).

[92] Rahmat Rahmat, Rob Kroeger, and Andrea Giammanco. "The fast simulation of the CMS experiment." In: *J. Phys. Conf. Ser.* 396 (2012), p. 062016. DOI: 10.1088/1742-6596/396/6/062016 (cit. on p. 40).

[93] A. M. Sirunyan et al. "Particle-flow reconstruction and global event description with the CMS detector." In: *JINST* 12.10 (2017), P10003. DOI: 10.1088/1748-0221/12/10/P10003. arXiv: 1706.04965 [physics.ins-det] (cit. on pp. 43, 44).

[94] T Speer, W Adam, R Frühwirth, A Strandlie, T Todorov, and M Winkler. *Track Reconstruction in the CMS Tracker*. Tech. rep. CMS-CR-2005-014. Geneva: CERN, 2005. URL: https://cds.cern.ch/record/884424 (cit. on p. 44).

[95] Vardan Khachatryan et al. "Performance of Electron Reconstruction and Selection with the CMS Detector in Proton-Proton Collisions at $\sqrt{s} = 8$ TeV." In: *JINST* 10.06 (2015), P06005. DOI: 10.1088/1748-0221/10/06/P06005. arXiv: 1502.02701 [physics.ins-det] (cit. on pp. 45, 47, 57, 160).

[96] CMS Collaboration. "Pileup Removal Algorithms." In: (2014) (cit. on p. 46).

[97] A. M. Sirunyan et al. "Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s} = 13$ TeV." In: *JINST* 13.06 (2018), P06015. DOI: 10.1088/1748-0221/13/06/P06015. arXiv: 1804.04528 [physics.ins-det] (cit. on pp. 47, 57).

[98] Philipp Pigard and Claude Charlot. "Electron studies and search for vector boson scattering in events with four leptons and two jets with the CMS detector at the LHC. Identification des électrons et mise en évidence de la diffusion de bosons massifs dans les évènements à quatre leptons et deux jets avec le détecteur CMS auprès du LHC." Presented 2017. 2017. URL: https://cds.cern.ch/record/2290139 (cit. on pp. 48, 49).

[99] "Commissioning of the Particle-Flow reconstruction in Minimum-Bias and Jet Events from pp Collisions at 7 TeV." In: (2010) (cit. on p. 49).

[100] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. "FastJet User Manual." In: *Eur. Phys. J.* C72 (2012), p. 1896. DOI: 10.1140/epjc/s10052-012-1896-2. arXiv: 1111.6097 [hep-ph] (cit. on p. 49).

[101] CMS Collaboration. "SUSY searches with multileptons." In: *CMS internal note* AN-2015/133 (2016). URL: http://cms.cern.ch/iCMS/jsp/openfile.jsp?tp=draft&files=AN2015_133_v12.pdf (cit. on p. 49).

[102] CMS Collaboration. "Measurement of the inclusive and differential cross section of top pair-production in association with a Z boson at 13 TeV." In: *CMS internal note* AN-2018/025 (2018). URL: http://cms.cern.ch/iCMS/jsp/openfile.jsp?tp=draft&files=AN2018_025_v8.pdf (cit. on p. 49).

[103] Vardan Khachatryan et al. "Search for new physics in same-sign dilepton events in proton-proton collisions at $\sqrt{s} = 13TeV$." In: *Eur. Phys. J. C* 76 (2016), p. 439. DOI: 10.1140/epjc/s10052-016-4261-z. arXiv: 1605.03171 [hep-ex] (cit. on p. 49).

[104] Andreas Hocker et al. "TMVA - Toolkit for Multivariate Data Analysis." In: (2007). arXiv: physics/0703039 [physics.data-an] (cit. on pp. 51, 163).

[105] CMS Collaboration. "Search for tZq in trilepton events." In: *CMS internal note* AN-2018/100 (2018). URL: http://cms.cern.ch/iCMS/jsp/openfile.jsp?tp=draft&files=AN2018_100_v6.pdf (cit. on p. 51).

[106] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. "The anti-k_t jet clustering algorithm." In: *JHEP* 04 (2008), p. 063. DOI: 10.1088/1126-6708/2008/04/063. arXiv: 0802.1189 [hep-ph] (cit. on pp. 53, 54).

[107] Gavin P. Salam. "Towards Jetography." In: *Eur. Phys. J.* C67 (2010), pp. 637–686. DOI: 10.1140/epjc/s10052-010-1314-6. arXiv: 0906.1833 [hep-ph] (cit. on p. 53).

[108] Serguei Chatrchyan et al. "Determination of Jet Energy Calibration and Transverse Momentum Resolution in CMS." In: *JINST* 6 (2011), P11002. DOI: 10.1088/1748-0221/6/11/P11002. arXiv: 1107.4277 [physics.ins-det] (cit. on p. 54).

[109] Vardan Khachatryan et al. "Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV." In: *JINST* 12.02 (2017), P02014. DOI: 10.1088/1748-0221/12/02/P02014. arXiv: 1607.03663 [hep-ex] (cit. on p. 54).

[110] Nicola Cabibbo. "Unitary Symmetry and Leptonic Decays." In: *Phys. Rev. Lett.* 10 (1963). [,648(1963)], pp. 531–533. DOI: 10.1103/PhysRevLett.10.531 (cit. on p. 55).

[111] Makoto Kobayashi and Toshihide Maskawa. "CP Violation in the Renormalizable Theory of Weak Interaction." In: *Prog. Theor. Phys.* 49 (1973), pp. 652–657. DOI: 10.1143/PTP.49.652 (cit. on p. 55).

[112] A. M. Sirunyan et al. "Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV." In: *JINST* 13.05 (2018), P05011. DOI: 10.1088/1748-0221/13/05/P05011. arXiv: 1712.07158 [physics.ins-det] (cit. on pp. 55, 56).

[113] Daniel Guest, Julian Collado, Pierre Baldi, Shih-Chieh Hsu, Gregor Urban, and Daniel Whiteson. "Jet Flavor Classification in High-Energy Physics with Deep Neural Networks." In: *Phys. Rev.* D94.11 (2016), p. 112002. DOI: 10.1103/PhysRevD.94.112002. arXiv: 1607.08633 [hep-ex] (cit. on p. 55).

[114] Vardan Khachatryan et al. "Performance of the CMS missing transverse momentum reconstruction in pp data at $\sqrt{s}$ = 8 TeV." In: *JINST* 10.02 (2015), P02006. DOI: 10.1088/1748-0221/10/02/P02006. arXiv: 1411.0511 [physics.ins-det] (cit. on p. 56).

[115] Georges Aad et al. "Search for supersymmetry at $\sqrt{s}$=8 TeV in final states with jets and two same-sign leptons or three leptons with the ATLAS detector." In: *JHEP* 06 (2014), p. 035. DOI: 10.1007/JHEP06(2014)035. arXiv: 1404.2500 [hep-ex] (cit. on p. 61).

[116] Serguei Chatrchyan et al. "Search for anomalous production of events with three or more leptons in *pp* collisions at $\sqrt{(s)}$ = 8 TeV." In: *Phys. Rev.* D90 (2014), p. 032006. DOI: 10.1103/PhysRevD.90.032006. arXiv: 1404.5801 [hep-ex] (cit. on p. 61).

[117] Vardan Khachatryan et al. "Searches for supersymmetry based on events with b jets and four W bosons in pp collisions at 8 TeV." In: *Phys. Lett.* B745 (2015), pp. 5–28. DOI: 10.1016/j.physletb.2015.04.002. arXiv: 1412.4109 [hep-ex] (cit. on p. 61).

[118] Vardan Khachatryan et al. "Search for new phenomena with multiple charged leptons in proton–proton collisions at $\sqrt{s}$ = 13 TeV." In: *Eur. Phys. J.* C77.9 (2017), p. 635. DOI: 10.1140/epjc/s10052-017-5182-1. arXiv: 1701.06940 [hep-ex] (cit. on pp. 61, 116).

[119] Jan Hendrik Hoss, Rainer Wallny, and Roland Horisberger. "Search for Supersymmetry with Multiple Charged Leptons at sqrts = 13 TeV with CMS and Radiation Tolerance of the Readout Chip for the Phase I Upgrade of the Pixel Detector." 2017. URL: https://cds.cern.ch/record/2287141 (cit. on p. 101).

[120] Richard D. Ball et al. "Parton distributions for the LHC Run II." In: *JHEP* 04 (2015), p. 040. DOI: 10.1007/JHEP04(2015)040. arXiv: 1410.8849 [hep-ph] (cit. on pp. 110, 248, 251, 253).

[121] Jon Butterworth et al. "PDF4LHC recommendations for LHC Run II." In: *J. Phys.* G43 (2016), p. 023001. DOI: 10.1088/0954-3899/43/2/023001. arXiv: 1510.03865 [hep-ph] (cit. on p. 110).

[122] *Procedure for the LHC Higgs boson search combination in Summer 2011*. Tech. rep. CMS-NOTE-2011-005. ATL-PHYS-PUB-2011-11. Geneva: CERN, 2011. URL: https://cds.cern.ch/record/1379837 (cit. on pp. 112, 171).

[123] Glen Cowan, Kyle Cranmer, Eilam Gross, and Ofer Vitells. "Asymptotic formulae for likelihood-based tests of new physics." In: *Eur. Phys. J.* C71 (2011). [Erratum: Eur. Phys. J.C73,2501(2013)], p. 1554. DOI: 10.1140/epjc/s10052-011-1554-0,10.1140/epjc/s10052-013-2501-z. arXiv: 1007.1727 [physics.data-an] (cit. on pp. 115, 172).

[124] Vardan Khachatryan et al. "Search for top-squark pairs decaying into Higgs or Z bosons in pp collisions at $\sqrt{s}$=8 TeV." In: *Phys. Lett.* B736 (2014), pp. 371–397. DOI: 10.1016/j.physletb.2014.07.053. arXiv: 1405.3886 [hep-ex] (cit. on p. 116).

[125] CMS. *CMS Supersymmetry Physics Results*. https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsSUS (cit. on p. 122).

[126] Morad Aaboud et al. "Search for direct top squark pair production in events with a Higgs or Z boson, and missing transverse momentum in $\sqrt{s} = 13$ TeV pp collisions with the ATLAS detector." In: *JHEP* 08 (2017), p. 006. DOI: 10.1007/JHEP08(2017)006. arXiv: 1706.03986 [hep-ex] (cit. on p. 116).

[127] Morad Aaboud et al. "Search for supersymmetry in final states with two same-sign or three leptons and jets using 36 fb^{-1} of $\sqrt{s} = 13$ TeV pp collision data with the ATLAS detector." In: *JHEP* 09 (2017), p. 084. DOI: 10.1007/JHEP09(2017)084. arXiv: 1706.03731 [hep-ex] (cit. on p. 116).

[128] Vardan Khachatryan et al. "Observation of top quark pairs produced in association with a vector boson in pp collisions at $\sqrt{s} = 8$ TeV." In: *JHEP* 01 (2016), p. 096. DOI: 10.1007/JHEP01(2016)096. arXiv: 1510.01131 [hep-ex] (cit. on p. 125).

[129] D. de Florian et al. "Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector." In: (2016). DOI: 10.23731/CYRM-2017-002. arXiv: 1610.07922 [hep-ph] (cit. on pp. 170, 174, 177, 217, 219).

[130] Deniz Poyraz, Ghent University Dobur Didar, and Ghent University Tytgat Michael. "Measurement of the top pair-production in association with a Z boson at LHC and WZ scattering in view of the High Luminosity LHC." Presented 19 Jun 2018. 2018. URL: https://cds.cern.ch/record/2635963 (cit. on p. 177).

[131] Peter Skands, Stefano Carrazza, and Juan Rojo. "Tuning PYTHIA 8.1: the Monash 2013 Tune." In: *Eur. Phys. J.* C74.8 (2014), p. 3024. DOI: 10.1140/epjc/s10052-014-3024-y. arXiv: 1404.5630 [hep-ph] (cit. on pp. 217, 248, 251, 253).

[132] Spyros Argyropoulos and Torbörn Sjöstrand. "Effects of color reconnection on $t\bar{t}$ final states at the LHC." In: *JHEP* 11 (2014), p. 043. DOI: 10.1007/JHEP11(2014)043. arXiv: 1407.6653 [hep-ph] (cit. on p. 217).

[133] Jesper R. Christiansen and Peter Z. Skands. "String formation beyond leading colour." In: *JHEP* 08 (2015), p. 003. DOI: 10.1007/JHEP08(2015)003. arXiv: 1505.01681 [hep-ph] (cit. on p. 217).

[134] Dao Thi Nhung, Le Duc Ninh, and Marcus M. Weber. "NLO corrections to WWZ production at the LHC." In: *JHEP* 12 (2013), p. 096. DOI: 10.1007/JHEP12(2013)096. arXiv: 1307.7403 [hep-ph] (cit. on p. 217).

[135] Morad Aaboud et al. "Measurement of the $t\bar{t}Z$ and $t\bar{t}W$ cross sections in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector." In: (2019). arXiv: 1901.03584 [hep-ex] (cit. on p. 218).

[136] Pierre Baldi, Kyle Cranmer, Taylor Faucett, Peter Sadowski, and Daniel Whiteson. "Parameterized neural networks for high-energy physics." In: *Eur. Phys. J.* C76.5 (2016), p. 235. DOI: 10.1140/epjc/s10052-016-4099-4. arXiv: 1601.07913 [hep-ex] (cit. on p. 224).

[137] Richard D. Ball et al. "Parton distributions from high-precision collider data." In: *Eur. Phys. J.* C77.10 (2017), p. 663. DOI: 10.1140/epjc/s10052-017-5199-5. arXiv: 1706.00428 [hep-ph] (cit. on pp. 248, 251, 253).

[138] Vardan Khachatryan et al. "Event generator tunes obtained from underlying event and multiparton scattering measurements." In: *Eur. Phys. J.* C76.3 (2016), p. 155. DOI: 10.1140/epjc/s10052-016-3988-x. arXiv: 1512.00815 [hep-ex] (cit. on pp. 248, 251, 253).

[139] CMS Collaboration. "Investigations of the impact of the parton shower tuning in Pythia 8 in the modelling of $t\bar{t}$ at $\sqrt{s} = 8$ and 13 TeV." In: (2016) (cit. on pp. 248, 251, 253).

[140] CMS Collaboration. "Extraction and validation of a new set of CMS PYTHIA8 tunes from underlying-event measurements." In: (2018) (cit. on p. 253).

[141] John M. Campbell and R. K. Ellis. "MCFM for the Tevatron and the LHC." In: *Nucl. Phys. Proc. Suppl.* 205-206 (2010), pp. 10–15. DOI: 10.1016/j.nuclphysbps.2010.08.011. arXiv: 1007.3492 [hep-ph] (cit. on p. 254).